

Review

A Comprehensive Survey of Deep Learning Approaches in Image Processing

Maria Trigka  and Elias Dritsas * Industrial Systems Institute (ISI), Athena Research and Innovation Center, 26504 Patras, Greece;
trigka@isi.gr

* Correspondence: dritsas@isi.gr

Abstract: The integration of deep learning (DL) into image processing has driven transformative advancements, enabling capabilities far beyond the reach of traditional methodologies. This survey offers an in-depth exploration of the DL approaches that have redefined image processing, tracing their evolution from early innovations to the latest state-of-the-art developments. It also analyzes the progression of architectural designs and learning paradigms that have significantly enhanced the ability to process and interpret complex visual data. Key advancements, such as techniques improving model efficiency, generalization, and robustness, are examined, showcasing DL's ability to address increasingly sophisticated image-processing tasks across diverse domains. Metrics used for rigorous model evaluation are also discussed, underscoring the importance of performance assessment in varied application contexts. The impact of DL in image processing is highlighted through its ability to tackle complex challenges and generate actionable insights. Finally, this survey identifies potential future directions, including the integration of emerging technologies like quantum computing and neuromorphic architectures for enhanced efficiency and federated learning for privacy-preserving training. Additionally, it highlights the potential of combining DL with emerging technologies such as edge computing and explainable artificial intelligence (AI) to address scalability and interpretability challenges. These advancements are positioned to further extend the capabilities and applications of DL, driving innovation in image processing.

Academic Editors: Chih-Chang Yu,
Jian-Jiun Ding and Feng-Tsun Chien

Received: 20 December 2024

Revised: 13 January 2025

Accepted: 13 January 2025

Published: 17 January 2025

Citation: Trigka, M.; Dritsas, E. A Comprehensive Survey of Deep Learning Approaches in Image Processing. *Sensors* **2025**, *25*, 531.
<https://doi.org/10.3390/s25020531>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: image processing; deep learning; techniques; models; metrics

1. Introduction

The field of image processing has been revolutionized by the advent of DL, a subset of AI inspired by the structure and processes of the human brain to analyze and interpret complex data patterns. Traditionally, image processing relied heavily on manual feature extraction and classical machine learning (ML) techniques, which required significant domain expertise and often struggled with the variability and complexity inherent in visual data. These methods, while effective in specific, well-defined tasks, lacked the flexibility and scalability needed to handle the diverse and high-dimensional nature of real-world images [1–3].

DL, characterized by its ability to learn hierarchical representations directly from raw data, has addressed many of the limitations of traditional approaches. The introduction of multi-layered neural networks (NNs) enabled models to automatically discover intricate patterns and features that were previously unachievable with manual techniques. This shift from handcrafted feature engineering to automated feature learning marked a pivotal

moment in image processing, allowing for significant advancements in both accuracy and generalizability across a broad range of applications [4–6].

One of the most significant breakthroughs in DL was the ability to process large-scale image datasets, which provided the foundation for developing robust and generalizable models. These models not only excelled in traditional image processing tasks, such as classification and segmentation, but also opened new avenues for innovation in areas that were previously considered too challenging or computationally prohibitive. The availability of large datasets and the increase in computational power, particularly through the use of Graphics Processing Units (GPUs), further accelerated this progress, making DL the dominant paradigm in image processing [7–9].

The architectural advancements in DL models have also played a crucial role in this evolution. The design of more complex and deeper networks, capable of capturing a wide range of visual features across different scales, has enabled the processing of images with unprecedented accuracy. These models have evolved to handle various aspects of image processing, from low-level tasks like denoising and super-resolution to high-level tasks such as object detection and semantic segmentation. Each new generation of models has built upon the successes of its predecessors, incorporating novel mechanisms to enhance learning efficiency, reduce computational costs, and improve model interpretability [10–12].

Moreover, the versatility of DL has facilitated its application across numerous domains, demonstrating its ability to solve complex and domain-specific challenges. The adaptability of DL models to different types of visual data—from natural images to medical scans—has led to breakthroughs in diverse fields, significantly impacting research and industry practices alike. This has established DL not just as a tool for solving image processing problems but as a fundamental technology driving innovation across a wide spectrum of scientific and technological endeavours [13–15].

Despite these advancements, the application of DL in image processing is not without challenges. The reliance on large, labelled datasets raises concerns about the scalability of these models to tasks where annotated data are scarce or difficult to obtain. Additionally, the high computational demands of training deep networks, particularly as models grow in complexity, pose significant barriers to entry for many researchers and practitioners. The interpretability of DL models also remains a critical issue, especially in high-stakes applications wherein understanding the model's decision-making process is as important as its accuracy [16–18].

The rapid advancements and widespread adoption of DL in image processing have led to an explosion of research, resulting in a vast and fragmented body of knowledge. As new models and techniques continuously emerge, it becomes increasingly challenging for researchers and practitioners to stay abreast of the latest developments and to understand how these advancements interrelate. This survey is motivated by the need to consolidate and synthesize this growing body of work, providing a structured and comprehensive overview that can serve as both a reference for current researchers and a guide for future exploration. Furthermore, while many reviews focus on specific aspects of DL in image processing—such as particular models or applications—there is a need for a broader survey that not only covers the evolution of models but also delves into the underlying techniques, evaluation metrics, and emerging trends. By addressing these elements, this survey aims to bridge the gap between theory and practice, offering insights that are relevant across a range of applications and research contexts. Specifically, this survey makes several key contributions to the field of DL in image processing:

- We provide an in-depth examination of the evolution of DL models in image processing, from foundational architectures to the latest advancements, highlighting the key developments that have shaped the field.

- The survey synthesizes various DL techniques that have been instrumental in advancing image processing, including those that enhance model efficiency, generalization, and robustness.
- We discuss the critical metrics used to evaluate DL models in image processing, offering a nuanced understanding of how these metrics are applied across different tasks.
- This survey identifies the persistent challenges in the application of DL to image processing and also explores potential future directions, including the integration of emerging technologies that could further advance the field.

The remaining paper is illustrated in Figure 1 and is structured as follows. Section 2 provides the evolution of DL in image processing. Moreover, Section 3 describes DL techniques in image processing. Section 4 notes advanced DL models. Next, in Section 5, the evaluation metrics for image processing models are provided. Section 6 presents applications of DL in image processing. Moreover, Section 7 illustrates challenges and future directions. Finally, Section 8 concludes the present survey.

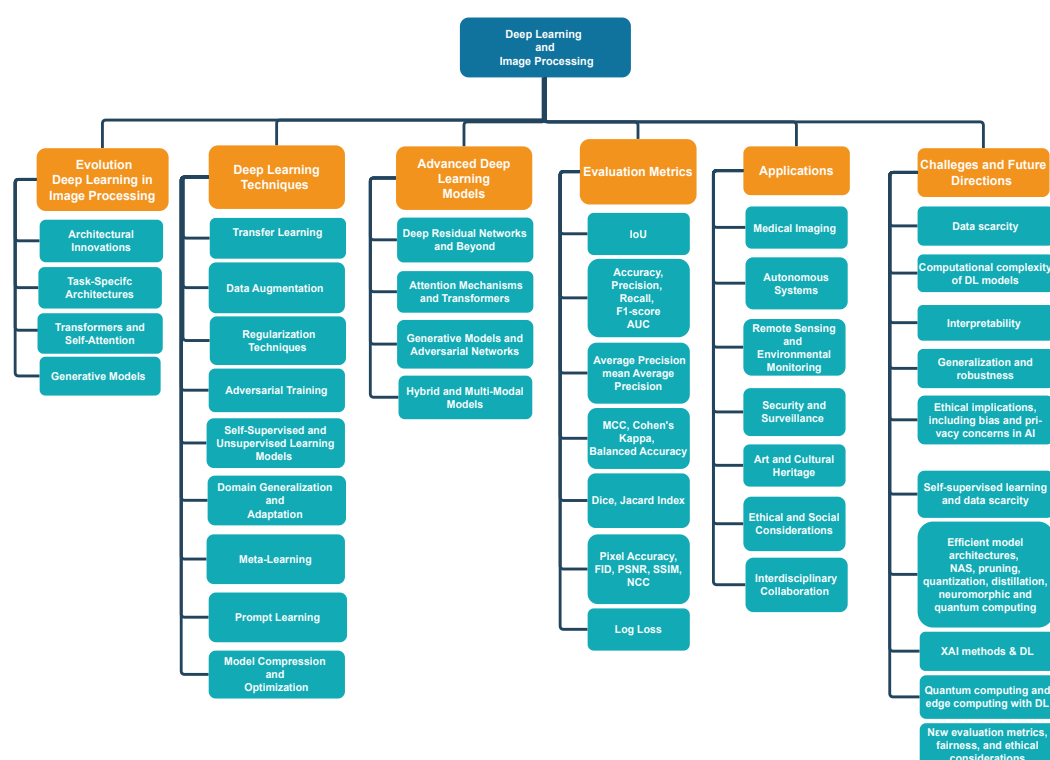


Figure 1. An overview of surveyed key topics in image processing with DL.

2. Evolution of Deep Learning in Image Processing

The evolution of DL in image processing represents a transformative journey from rudimentary NN models to modern architectures capable of handling complex visual data with unparalleled accuracy. This evolution is marked by several pivotal developments that have redefined the capabilities of image processing systems, pushing the boundaries of what was once thought possible in the field.

2.1. Architectural Innovations

The architectural evolution in DL for image processing has been pivotal in addressing the complex challenges posed by high-dimensional visual data. Convolutional NNs (CNNs) laid the groundwork by effectively capturing spatial hierarchies through convolutional layers [19]. However, the introduction of deeper networks, such as residual networks (ResNets), marked a significant leap forward. ResNets leverage residual connections to

bypass one or more layers, mitigating the vanishing gradient problem and enabling the training of exceptionally deep networks. This advancement allows these models to learn richer and more intricate features, leading to substantial improvements in tasks such as image classification and object detection [20,21]. Furthermore, the densely connected convolutional network (DenseNet), with its densely connected layers, further enhances this capability by promoting feature reuse across layers, reducing the number of parameters required, and improving both computational efficiency and model accuracy [22–24].

Multi-branch architectures, exemplified by inception networks, represent another significant development, enabling models to capture information at multiple scales within the same architecture. This design allows the network to process various feature scales simultaneously, enhancing its ability to generalize across different image-processing tasks. Such architectures are particularly effective in handling the diverse and complex nature of visual data, making them ideal for advanced tasks like semantic segmentation and image synthesis. The integration of these architectural innovations has not only pushed the boundaries of what DL models can achieve but also set new standards for performance in the field of image processing [25–28].

The field of object detection has seen remarkable progress with the introduction of models like YOLO (You Only Look Once), which revolutionized real-time detection by using a single NN to predict bounding boxes and class probabilities simultaneously. Unlike traditional methods that rely on region proposals, YOLO's unified approach significantly reduces computational complexity while maintaining accuracy, making it a preferred choice for applications requiring speed and efficiency. Its ongoing development, from YOLOv1 to YOLOv8, demonstrates its adaptability and continued relevance in DL research [29,30].

The next generation of convolutional networks (ConvNext) is a modernized CNN that integrates design principles from vision transformers while retaining the simplicity and efficiency of traditional CNNs. It revisits standard convolutional architectures and improves them with innovations like depth-wise convolutions, layer normalization, and expanded kernel sizes, achieving competitive performance in image classification, object detection, and segmentation. ConvNext bridges the gap between CNNs and attention-based architectures, combining the strengths of both approaches [31–33].

2.2. Specialized Architectures for Task-Specific Challenges

As DL models evolved, the need for specialized architectures tailored to specific image-processing tasks became apparent. Fully convolutional networks (FCNs) and U-Net architectures were developed to address the challenges of pixel-level predictions required in semantic segmentation. FCNs replace fully connected layers with convolutional layers, maintaining spatial hierarchies and enabling dense prediction tasks [34]. U-Net, with its encoder–decoder structure and skip connections, is particularly effective in capturing both contextual information and fine-grained details [35]. These features make U-Net highly suitable for medical imaging and other applications wherein precise boundary delineation is critical [36,37].

In object detection, mask region-based convolutional NNs (R-CNN) have set a new benchmark by extending the capabilities of region-based CNNs to include pixel-level segmentation [38]. This architecture integrates detection and segmentation tasks within a unified framework, enabling comprehensive scene understanding. The ability to generate high-quality segmentation masks for detected objects has proven invaluable in applications requiring detailed scene analysis, such as autonomous driving and video surveillance [39]. The development of these specialized architectures underscores the importance of designing task-specific solutions to meet the growing demands of advanced image processing challenges [40–42].

2.3. Expanding Capabilities with Transformers and Self-Attention

The introduction of self-attention mechanisms, particularly through vision transformers (ViTs), has expanded the capabilities of DL models in image processing. Unlike traditional CNNs, which focus on local features through fixed convolutional filters, transformers model global dependencies within an image. This capability allows ViTs to capture long-range relationships that are crucial for understanding complex scenes, making them particularly effective in tasks that require holistic image analysis, such as scene segmentation and object recognition. The scalability of transformers, which can be achieved with minimal architectural changes, makes them well-suited for handling large and diverse datasets [43–47].

Self-attention mechanisms have also paved the way for more flexible and powerful models. By dynamically focusing on different parts of an image based on task relevance, these models can prioritize critical features while ignoring irrelevant data. This selective attention mechanism enhances the model's ability to generalize across varied image-processing tasks. As transformers continue to evolve, their integration into hybrid architectures that combine the strengths of both CNNs and transformers is likely to yield even greater performance gains, further pushing the boundaries of what DL models can achieve in the field of image processing [48–53].

2.4. Integration of Generative Models

Generative models, particularly generative adversarial networks (GANs), have introduced new dimensions to DL in image processing. GANs have revolutionized the field by enabling the generation of high-quality, realistic images through a competitive training process involving two networks: the generator and the discriminator. This framework allows GANs to learn complex data distributions without explicit probabilistic modeling, making them highly effective for tasks such as image synthesis, style transfer, and super-resolution [54–57]. Advanced variants of GANs, like conditional GANs (CGANs) [58] and Wasserstein GANs (WGANs) [59], have further refined the generative process, addressing challenges such as mode collapse and ensuring more stable training.

Beyond their generative capabilities, GANs have significantly impacted other areas of image processing, such as data augmentation and domain adaptation. In scenarios where labeled data are scarce, GANs can generate synthetic data that closely resemble real-world samples, improving model robustness and generalization. Additionally, GANs are used in domain adaptation to align feature distributions between different domains, facilitating the transfer of models across diverse imaging contexts. The versatility and effectiveness of GANs in enhancing image-processing tasks underscore their importance as a core component of modern DL frameworks [60–64].

Finally, diffusion models have emerged as state-of-the-art generative models, excelling in image synthesis, denoising, and restoration by employing a probabilistic framework to reconstruct data from noise. These models work by gradually adding random noise to data during the forward process and then learning to reverse this process to generate high-quality outputs. This unique approach allows diffusion models to produce highly realistic and diverse data, often surpassing traditional GANs in terms of stability and output quality [65,66].

As illustrated in Table 1, the evolution of DL architectures in image processing has been marked by pivotal innovations. This table summarizes key references that outline the advancements in architectural innovations, scale-aware networks, and task-specific designs, highlighting the impact of DL on various image-processing challenges. The references serve as a foundation for understanding the significant strides made in the field, from foundational models to specialized architectures tailored for complex tasks.

Table 1. Summary of architectural innovations in DL for image processing.

Architecture	Innovation	References
CNNs	They are the foundation for image processing, enable automatic spatial hierarchy capture through convolutional layers, which process image patterns at different levels of granularity.	[19]
ResNets	Introduced residual connections to address vanishing gradient problems, allowing deeper networks to be trained by learning residual functions rather than direct mappings.	[20,21]
DenseNets	DenseNets enable direct connections between all layers to enhance feature reuse, reduce computational costs, and improve efficiency in image classification and object detection tasks.	[22–24]
Multi-branch Architectures	Inception networks process image features at multiple scales simultaneously within a single model, significantly improving performance on complex tasks like semantic segmentation.	[25–28]
YOLO	YOLO transformed object detection with a single network approach, simultaneously predicting bounding boxes and class probabilities, achieving real-time efficiency with high accuracy.	[29,30]
ConvNext	ConvNext integrates principles from vision transformers into traditional CNNs, improving performance with innovations like depth-wise convolutions and larger kernel sizes while retaining simplicity.	[31–33]
FCNs	FCNs replace fully connected layers with convolutional ones, preserving spatial hierarchies for dense predictions in tasks such as semantic segmentation.	[34]
U-Net	U-Net’s encoder–decoder structure with skip connections enables precise boundary delineation, making it especially effective for medical imaging and other pixel-level prediction tasks.	[35–37]
Mask R-CNN	Mask R-CNN extends object detection by integrating segmentation, creating pixel-level masks for detected objects, which is valuable for tasks like autonomous driving and video analysis.	[38,39]
Specialized Task-Specific Architectures	Tailored architectures address specific challenges in advanced image processing, ensuring accuracy and efficiency in highly specialized domains.	[40–42]
ViTs	Vision transformers handle global dependencies in images by modeling them as sequences of patches, offering advantages in scene understanding and holistic image analysis	[43–47]
Self-Attention Mechanisms	Self-attention dynamically prioritizes relevant image regions for tasks like classification and segmentation, enabling robust generalization across diverse datasets.	[48–53]
GANs	GANs use adversarial training between a generator and discriminator to create realistic images, excelling in tasks like image synthesis, super-resolution, and style transfer.	[54–57]
CGANs	CGANs integrate class labels or other auxiliary information into GANs, enabling controlled generation of specific types of images based on given conditions.	[58]
WGANs	WGANs improve GAN training stability by introducing a Wasserstein distance-based loss function, addressing mode collapse and convergence issues.	[59]
Other GAN Applications	GANs are used for synthetic data generation, data augmentation, and domain adaptation, improving robustness and generalization in low-data scenarios and cross-domain tasks.	[60–64]
Diffusion Models	Diffusion models utilize a probabilistic framework to iteratively add and remove noise, achieving state-of-the-art results in tasks like image restoration, synthesis, and denoising.	[65,66]

3. Deep Learning Techniques in Image Processing

The rapid advancements in DL have been driven not only by the evolution of NN architectures but also by the development of sophisticated techniques that optimize these

models for specific image-processing tasks. These techniques are critical to enhancing the capabilities of DL models in terms of accuracy, efficiency, and generalization. This section explores several key techniques that have profoundly impacted the field.

3.1. Transfer Learning

Transfer learning has become a cornerstone in DL, particularly within image processing, where it addresses the challenges of training models on limited datasets by leveraging pre-trained models on large, diverse datasets. The principle behind transfer learning is to utilize the feature representations learned from a source task and apply them to a target task, effectively reducing the need for extensive labelled data and computational resources.

According to [67], deep transfer learning approaches can be categorized into two main types: adversarial-based and network-based methods. Adversarial-based techniques leverage adversarial learning strategies to enhance performance. On the other hand, network-based approaches include fine-tuning, freezing CNN layers, and progressive learning, enabling the adaptation of pre-trained models to new tasks by varying the level of layer adjustments and optimization.

The concept of transfer learning is particularly valuable in domains like medical imaging [68], where acquiring large labelled datasets is costly or impractical. By fine-tuning pre-trained models on smaller, task-specific datasets, practitioners can achieve significant performance improvements, often surpassing models trained from scratch [69,70]. The authors in [71] emphasized the application of transformer-based pre-trained models for image processing tasks such as super-resolution and denoising. Leveraging transformers' ability to capture global dependencies, and adapt to task-specific needs, trained on a large-scale synthetic dataset, the model demonstrated exceptional generalization across diverse tasks. Moreover, the combination of transfer learning with pre-trained CNNs [72] especially in medical image classification is analyzed in [73], focusing on approaches like feature extraction and fine-tuning. It provides actionable insights for leveraging transfer learning to address data scarcity, including recommendations on optimal model selection and configuration. Previous works highlight the transformative potential of pre-trained models in various image-processing tasks.

However, while transfer learning offers many benefits, it is not without challenges. One such challenge is negative transfer, where the knowledge from the pre-trained model may not always be beneficial and could even hinder performance on the target task. This occurs when the source and target tasks are dissimilar, leading to ineffective feature reuse [74,75]. Several categories and approaches to mitigate negative transfer in transfer learning are detailed in [76]. These categories include data transferability enhancement, with methods like domain-level, instance-level, feature-level, and class-level strategies. Model transferability enhancement focuses on techniques like transferable batch normalization, adversarial training, multiple models, parameter selection, and parameter regularization. Training process enhancement involves hyper-parameter tuning and gradient correction. Target prediction enhancement includes soft labeling, selective labeling, weighted clustering, and entropy regularization. Moreover, concept-wise fine-tuning was presented in [77], addressing the problem by maximizing mutual information for rare features and applying causal adjustment to correct spurious correlations, enhancing transfer robustness and effectiveness. Concept-wise fine-tuning falls under model transferability enhancement, emphasizing parameter selection and regularization. These approaches collectively enhance transfer robustness and minimize the impact of irrelevant or harmful knowledge.

A recent study in [78] suggested an approach based on hierarchical transfer progressive learning (HTPL), demonstrating its effectiveness in addressing negative transfer by

progressively fine-tuning knowledge from source to target tasks. It begins with transferring general low-level features, which are less domain-specific, before gradually incorporating high-level features tailored to sonar image characteristics such as low resolution and speckle noise. This staged adaptation minimizes the risk of transferring irrelevant knowledge, ensuring effective domain alignment. Apart from its contribution to negative transfer mitigation, it also enhances fine-grained feature extraction, addresses the scarcity of labeled data with self-supervised pre-training, and resolves class imbalance using key point sensitive loss. These strategies demonstrate the solution's robustness in sonar image classification challenges. A summary of topics discussed regarding transfer learning techniques is presented in Table 2.

Table 2. Summary of topics in the context of transfer learning.

Topic	Description	References
Pre-trained models and transfer learning strategies	Analyze transfer learning using CNN- and transformer-based pre-trained models and their application in medical imaging. Also focused on key categories, i.e., adversarial-based and network-based (fine-tuning, freezing CNN layers, and progressive learning).	[67–73]
Negative transfer	Highlights the issue of negative transfer, where source and target tasks differ significantly, hindering performance, and strategies to mitigate its impact.	[74,75]
Negative transfer: mitigation strategies	Explores data transferability, model transferability, training process enhancement, and prediction refinement strategies.	[76]
	Concept-wise fine-tuning is a model transferability method.	[77]
	HTPL, a feature-based transfer learning approach that progressively fine-tunes features ensuring effective domain alignment and mitigating negative transfer issues.	[78]

3.2. Data Augmentation

Data augmentation is crucial in DL, especially in image processing, where the diversity and volume of training data significantly influence model performance. By systematically applying a series of transformations to the original dataset, data augmentation increases the effective size of the dataset and enhances the model's ability to generalize by exposing it to a broader range of variations and distortions. This technique is particularly important in preventing overfitting, especially when acquiring large datasets is impractical or expensive [79–83].

Advanced data augmentation strategies have evolved beyond basic geometric transformations like rotation and scaling [84]. Techniques such as Cutout [85], Mixup [86], and CutMix [87] introduce more complex variations by blending different image samples or masking out regions of images, encouraging the model to focus on global context rather than specific localized features. The advent of automated data augmentation methods, such as AutoAugment and RandAugment, represents a significant leap forward. These

methods use reinforcement learning (RL) and optimization algorithms to automatically discover the most effective augmentation strategies tailored to the specific dataset and task at hand. This reduces the manual effort involved and consistently results in superior model performance, particularly in complex image-processing tasks [88–90]. Table 3 provides a comprehensive summary of the techniques addressed on data augmentation.

Table 3. Summary of data augmentation techniques.

Technique	Description	References
Basic Methods	Basic augmentation techniques (e.g., rotation, scaling) for increasing dataset diversity and preventing overfitting, especially when large datasets are impractical or expensive.	[79–83]
Advanced Modern Approaches	Techniques like Cutout, Mixup, and CutMix enhance model robustness by introducing complex image variations and encouraging focus on global context rather than localized features.	[84]
Complex Image Transformations	Reviews the application of techniques such as blending images, masking, and targeted transformations to improve model generalization and reduce overfitting.	[85–87]
Automated Strategies	AutoAugment and RandAugment that leverage optimization and RL to identify the most effective augmentation policies for specific datasets, significantly improving performance with reduced manual effort.	[88–90]

3.3. Regularization Techniques

Regularization techniques are essential for DL models, particularly in image processing, where high dimensionality and complexity of visual data often lead to overfitting which occurs when a model learns not just the underlying patterns in the training data but also the noise and irrelevant details, resulting in poor generalization to unseen data. Various regularization techniques to address overfitting and enhance generalization in DL models are systematically reviewed in [91]. Also, it compares traditional and modern methods, evaluating their computational costs and impact on model performance, with experimental insights to guide practical applications. Key regularization strategies include dropout, weight decay, and batch normalization. Apart from the L_2 regularization (i.e., weight decay), it explores variations in dropout, like DropAll, Curriculum Dropout, and DropMaps to address co-adaptation and improve model robustness.

Dropout [92,93] is a widely used technique that randomly deactivates a subset of neurons during each training iteration, forcing the network to learn redundant representations of features. This reduces the risk of co-adaptation between neurons and enhances the network's robustness [94]. Weight decay [95] adds a penalty term to the loss function based on the magnitude of the network's weights, discouraging the model from assigning excessive importance to any particular weight and thus preventing overfitting. Ref. [96] explores disharmony issues between weight decay and weight normalization methods, offering insights into balancing these regularization strategies. Batch normalization [97], while primarily designed to stabilize and accelerate training, also functions as a regularization technique by reducing internal covariate shift, allowing for higher learning rates and improving overall model performance. The combination of these techniques ensures that models trained on complex image datasets are more likely to generalize well to new

data [98]. An overview of the covered techniques related to regularization is provided in Table 4.

Table 4. Summary of regularization topics and techniques.

Topic/Technique	Description	References
Comprehensive survey in Regularization	Reviews traditional and modern regularization methods, comparing their effectiveness, computational cost, and applicability to mitigate overfitting in DL.	[91]
Dropout	Prevents overfitting by randomly deactivating neurons during training, forcing the model to learn redundant feature representations and increasing robustness.	[92–94]
Weight Decay	Penalizes large weights by adding a regularization term to the loss function, discouraging excessive reliance on specific parameters and improving generalization. Also, disharmony issues are discussed with weight normalization methods.	[95,96]
Batch Normalization	Stabilizes and accelerates training by normalizing layer inputs, reducing internal covariate shift, and indirectly functioning as a regularization technique to improve model performance.	[97,98]

3.4. Adversarial Training

Adversarial training has emerged as a critical technique in DL, particularly for improving the robustness and security of models in image processing. This approach involves deliberately introducing adversarial examples—inputs subtly perturbed to deceive the model—into the training process. The goal is to fortify the model against potential vulnerabilities by exposing it to these adversarial inputs, thereby enhancing its resilience to attacks that could exploit weaknesses in its predictive capabilities [99–103].

However, generating effective adversarial examples that are both imperceptible to humans and capable of misleading the model remains a challenge. Techniques such as the Fast Gradient Sign Method (FGSM) [104] and Projected Gradient Descent (PGD) [105] have been developed to create these adversarial examples efficiently. These methods perturb input data minimally yet significantly affect the model’s output, challenging the model to learn more robust and invariant representations. Beyond improving robustness, adversarial training has broader implications, such as enhancing the model’s understanding of data distributions and improving generalization.

As research in adversarial training continues, it aims to strike an optimal balance between robustness and accuracy, making DL models more secure and reliable for real-world applications. Free adversarial training reduces the computational cost of adversarial training by reusing gradient computations through minibatch replays, updating both model weights and input perturbations simultaneously. It achieves adversarial robustness comparable to PGD training but with significantly fewer gradient calculations, making it computationally efficient [106]. Universal projected gradient descent (UPGD) is an enhanced adversarial attack method that generates universal perturbations effectively across multiple models and datasets by refining perturbations iteratively. It balances robustness and accuracy, achieving higher fooling rates and cross-model generalization compared to traditional techniques [107]. Finally, model-based adversarial training (MAT) extends traditional adversarial training by leveraging models of natural variation (such

as changes in lighting, weather, or resolution) to craft adversarial examples that simulate realistic shifts in data distribution. MAT is directly tied to adversarial training, bridging the gap between adversarial robustness and natural shift generalization [108]. A summarization of the methods in the topic of adversarial training is made in Table 5.

Table 5. Summary of aspects in adversarial training.

Topic	Description	References
Adversarial examples and Training	Overview of methods, challenges, and opportunities for generating adversarial examples to expose and improve model robustness.	[99–103]
FGSM	A computationally efficient method for generating adversarial examples with minimal perturbation.	[104]
PGD	An iterative approach and stronger method for crafting adversarial examples by refining perturbations stepwise.	[105]
Free Adversarial Training	Efficiently reuses gradient computations via minibatch replays to achieve robustness with reduced cost.	[106]
UPGD	Enhanced PGD algorithm for generating universal adversarial perturbations, balancing accuracy and robustness.	[107]
MAT	Leverages models of natural variation to generate adversarial examples, enhancing robustness against naturally shifted datasets.	[108]

3.5. Self-Supervised and Unsupervised Learning

Self-supervised and unsupervised learning has emerged as a transformative approach in DL, particularly for image processing tasks where labeled data are scarce or expensive to obtain [109]. Unlike traditional supervised learning, which relies on large, manually annotated datasets, self-supervised learning leverages vast amounts of unlabeled data by generating proxy tasks that can be solved without human intervention [110]. These proxy tasks enable the model to learn useful representations from the data, capturing images' underlying structure and semantics. The learned representations can then be fine-tuned for specific/downstream tasks such as classification, segmentation, or detection, often yielding performance that rivals fully supervised methods [111–116].

Recent innovations in self-supervised learning, such as contrastive learning, have shown remarkable success in this area [117]. Contrastive learning techniques like the simple framework for contrastive learning of visual (SimCLR) [118] and Momentum Contrast (MoCo) [119] enable models to learn robust representations by maximizing agreement between different augmentations of the same instance. This approach significantly reduces the dependency on labeled datasets, making DL more accessible and scalable across various domains. The integration of self-supervised learning with other advanced techniques, such as attention mechanisms, promises to further enhance the capabilities and applicability of DL in image processing [120–122].

Unsupervised learning, which focuses on uncovering the intrinsic structure of data without any explicit labels, has also seen significant advancements: in clustering using the Uniform Manifold Approximation and Projection (UMAP) dimensionality reduction technique [123], dimensionality reduction [124]. In addition to the previous advances, 3D convolutional autoencoders [125] have been instrumental in learning compact, latent repre-

sentations of images that preserve essential features while reducing noise and redundancy. The recent success of contrastive learning, a method that maximizes the similarity between different augmentations of the same image, has further pushed the boundaries of what can be achieved with minimal supervision. These approaches are particularly valuable in fields like medical imaging, where labeled data are limited, and are likely to become increasingly important as the field progresses [126–128]. An extended list of techniques and their purpose of use is concisely captured in Table 6.

Table 6. Summary of self-supervised and unsupervised learning techniques.

Technique	Description	References
Self-supervised and Unsupervised Learning	Techniques focusing on how they improve image classification without labeled data.	[109]
Proxy-based Learning	Frameworks for spectral–spatial hyperspectral image classification, enhancing robustness and efficiency.	[110]
Contrastive Learning	Reviews contrastive learning techniques in self-supervised frameworks, highlighting their success in extracting meaningful representations.	[111]
Multi-task Learning	Proposes frameworks for self-supervised learning in specific domains like medical imaging.	[112]
Self-supervised Learning Improvements	Mitigating issues related to proxy task specificity, improving performance across various downstream applications.	[113]
Representation Learning	Explores the combinations of labeled and unlabeled data for unified speech and visual representation.	[114]
	Reviews advances and challenges in self-supervised representation learning, emphasizing its potential in scalable applications.	[115]
	Context Autoencoders: demonstrates their use for effective representation learning in image processing tasks.	[116]
Self-supervised methods	Reviews key methods such as SimCLR and MoCo, focusing on robust representation learning through contrastive approaches.	[117–119]
Advanced self-supervised techniques	Integration of attention mechanisms and other advanced approaches to enhance capabilities in self-supervised frameworks.	[120–122]
Unsupervised Learning and Dimensionality reduction	Examines the use of clustering with UMAP in uncovering intrinsic structures of data, improving latent feature representations. Highlights the effectiveness of preserving essential features while reducing redundancy.	[123,124]
3D Convolutional Autoencoders	Explains their application in compact representation of hyperspectral image data.	[125]
Unsupervised Learning frameworks	Discusses applications into tasks such as medical image analysis, denoising, and segmentation.	[126–128]

3.6. Domain Generalization and Adaptation

In image processing, domain variability poses a significant challenge when models trained on one dataset fail to perform adequately on another due to differences in data distribution, a problem known as domain shifts. These shifts can arise from variations in lighting, resolution, imaging devices, or environmental conditions. For example, satellite images captured under varying weather conditions or medical images from scanners with differing configurations often exhibit discrepancies that hinder model generalization [129].

Domain generalization [130] aims to train models that perform robustly on unseen domains without direct access to their data during training. This is achieved by encouraging

models to learn domain-invariant features—representations that capture the essence of the data while disregarding domain-specific variations. Techniques such as deep domain confusion (DDC) [131] and domain-invariant component analysis (DICA) [132] are widely used to align feature distributions across multiple source domains. For instance, in medical imaging, a domain-generalized model trained on datasets from diverse hospitals can classify anomalies in scans from a new hospital, even if the imaging protocols differ. Episodic training frameworks, where models simulate potential domain shifts during training, further enhance robustness by preparing the model for unseen variations [133].

Domain adaptation, by contrast, assumes access to target domain data, albeit often unlabeled, during training. This enables explicit alignment of the source and target domain distributions to improve performance on the target domain. A popular approach is adversarial learning, implemented in models like domain-adversarial NNs (DANN) [134]. Here, a domain classifier guides the feature extractor in producing domain-agnostic representations, ensuring the model generalizes well to both domains. Another powerful tool is CycleGAN, which employs style transfer to transform target domain images into the appearance of the source domain [135]. For example, CycleGAN has been used to adapt object detection models for autonomous vehicles, enabling them to perform effectively in rural settings despite being trained in urban environments. The style transfer process aligns lighting, textures, and other visual properties between the domains, ensuring consistent detection accuracy [136]. In conclusion, Table 7 briefly presents the key techniques discussed in this section.

Table 7. Summary of domain generalization and adaptation techniques.

Technique	Description	References
Domain variability	Discusses the challenges caused by domain shifts, such as differences in lighting, resolution, or imaging devices.	[129]
Domain generalization	Focuses on training models to perform robustly on unseen domains without direct access to target domain data during training.	[130]
DDC	Aligns feature distributions across multiple source domains for robust generalization.	[131]
DICA	Aligns features for learning domain-invariant representations.	[132]
Episodic training	Frameworks to enhance robustness by simulating domain shifts during training to prepare models for unseen variations.	[133]
DANNs	Adversarial learning to generate domain-agnostic features, improving generalization across domains.	[134]
CycleGAN	Transform target domain images into the appearance of the source domain (style transfer), improving alignment.	[135]
	Demonstrates its application in autonomous driving for adapting object detection models in rural environments despite training in urban areas.	[136]

3.7. Meta-Learning

Meta-learning, or “learning to learn”, has emerged as a transformative approach in image processing, addressing the challenge of limited labeled data by enabling models to adapt quickly to new tasks. Unlike traditional DL, which requires extensive datasets, meta-learning trains models on diverse tasks to optimize their ability to generalize with minimal data [137].

There are three state-of-the-art types of meta-learning methods for image segmentation: metric-based, model-based, and optimization-based [138,139]. Metric-based approaches, such as prototypical/prototype networks, classify new data points by comparing them to learned class prototypes. Other methods in this category include siamese NNs and matching networks, which rely on feature extractors, similarity metrics, and automatic algorithm selection.

Model-based approaches aim to adapt to new tasks by changing the model’s learnable parameters. For example, memory-augmented NNs (MANNs) combine NNs with external memory modules to enhance learning efficiency. Despite their advantages, MANNs are complex, and meta-networks are computationally intensive with high memory requirements. Alternatively, the simple neural attentive meta-learner (SNAIL) offers a relatively straightforward structure but requires optimization for automatic parameter tuning and reduced computational demands [140].

Optimization-based approaches treat meta-learning as an optimization problem, aiming to extract meta-knowledge that improves optimization performance. These methods generate classifiers capable of performing well on a query set with only a few gradient updates. Model-agnostic meta-learning (MAML) is a widely used method in this category, fine-tuning model parameters for rapid adaptation. Other notable methods include META-LSTM and META-SGD, which leverage long short-term memory (LSTM) and stochastic gradient descent (SGD), respectively. Finally, Reptile, similar to MAML, adapts to new tasks by learning optimal initial parameters but is better suited for problems requiring numerous update steps. With lower variance, it achieves faster convergence but has primarily been validated for few-shot classification, with limited evidence for tasks like regression or RL [138]. To sum up, in Table 8, we capture the meta-learning categories and methods presented previously.

Table 8. Summary of meta-learning categories and methods.

Technique	Description	References
Metric-based	Prototypical, siamese, and matching networks classify new data points by comparing them to learned class prototypes using feature extractors and similarity metrics.	[137,138]
Model-based	MANNs combine NNs with external memory modules to enhance learning efficiency. The SNAIL improves parameter tuning efficiency.	[139]
Optimization-based	MAML fine-tunes model parameters for rapid task adaptation. Other methods: META-LSTM, META-SGD, Reptile.	[138–140]

3.8. Prompt Learning

Prompt learning is an emerging and impactful paradigm in image processing, enabling pre-trained models to adapt to specific tasks with minimal fine-tuning by embedding task-specific “prompts” into the input data [141]. Inspired by its success in natural language processing, prompt learning has demonstrated its utility in vision-language models like

Contrastive Language-Image Pretraining (CLIP) [142,143]. CLIP interprets textual prompts such as “a photo of a deforested area” or “a picture of an urban landscape” to classify images, making it highly versatile for applications like environmental monitoring and disaster assessment [144].

In interactive segmentation, click prompt learning allows for the real-time refinement of outputs. For example, interactive tools utilize user-provided prompts, such as clicks or bounding boxes, to guide segmentation tasks, a technique particularly valuable in medical imaging [145,146]. Also, a Promptable and Robust Interactive Segmentation Model (PRISM), with visual prompts aiming for precise segmentation of 3D medical images, is suggested in [147]. A systematic review and taxonomy of deep interactive segmentation of medical images is thoroughly described in [148], identifying key methods, models, and trends within the field while thoroughly discussing the related challenges. Additionally, prompt learning advances zero-shot learning, where task-specific prompts enable models trained on general datasets to tackle niche tasks without additional retraining. For instance, satellite imagery models can interpret prompts like “detect water bodies” to identify specific geographic features efficiently [149]. In Table 9, the key topics covered in this study concerning prompt learning are outlined.

Table 9. Summary of prompt learning techniques.

Technique	Description	References
Vision-language models (e.g., CLIP)	Enables classification tasks using textual prompts enhancing applications like environmental monitoring and disaster assessment.	[141–144]
Interactive segmentation	Click prompt learning for real-time refinement of outputs using user-provided prompts particularly useful in medical imaging. Interactive Medical Image Learning Framework using DL algorithms trained during the user study compared in performance against state-of-the-art modern augmentations. PRISM model applied to 3D medical images’ segmentation with significantly improved performance. In-depth analysis of the foundational principles of interactive segmentation methodologies and categorization based on common characteristics in the field of medical imaging.	[145–148]
Zero-shot learning	Utilizes task-specific prompts to adapt models trained on general datasets for niche tasks, such as land-use mapping or satellite imagery analysis without retraining.	[149]

3.9. Model Compression and Optimization Techniques for Efficiency and Scalability

As DL models become increasingly complex, model compression [150] and optimization techniques [151] have become essential for ensuring their efficiency and scalability, particularly in resource-constrained environments. Pruning is a key technique in model compression that reduces the size of a model by eliminating redundant or less significant parameters such as weights, neurons, or layers [152,153]. This not only decreases the model’s computational demands but also accelerates inference time and reduces memory usage, making it feasible to deploy DL models on edge devices with limited resources [154]. Advanced pruning strategies, such as those guided by RL, ensure that models retain their performance while becoming more efficient [155–157].

Quantization further contributes to optimization by reducing the precision of model parameters, converting them from 32-bit floating-point numbers to lower-bit representations, such as 8-bit integers [158,159]. This reduction significantly lowers the computational and memory requirements, enabling faster inference without compromising accuracy. Knowledge distillation is another powerful technique where a smaller, more efficient model (the student) learns from a larger, more accurate model (the teacher) [160]. This approach ensures that the student model retains the essential characteristics of the teacher while being optimized for deployment in real-time or resource-constrained environments. Together, these advanced optimization techniques are essential for extending the applicability of DL models across a wide range of real-world scenarios [161–167]. Finally, Table 10 categorizes the analyzed methods and highlights notable studies offering a reference for understanding the contributions of each technique to the field.

Table 10. Summary of model compression and optimization techniques.

Technique	Description	References
Model Compression Overview	Highlights the need for compression in DL to reduce computational cost and memory usage for deployment in resource-constrained environments such as edge devices.	[150]
Pruning	Removing redundancy or insignificant parameters (e.g., weights, neurons, layers) to reduce the model size, computational requirements, and inference time. Includes structured and unstructured pruning approaches.	[151–157]
Quantization	Reduced parameter precision (e.g., from 32-bit floats to 8-bit integers) for faster computation and memory usage while preserving model accuracy, enabling efficient deployment.	[158,159]
Knowledge Distillation	Transfers knowledge from a large teacher model to a smaller student model, preserving essential characteristics while enhancing scalability and efficiency for deployment.	[160,161]
Energy-Efficient Architectures	Focuses on developing architectures that consume less power and are optimized for specific hardware, including FPGA and ASIC implementations.	[162]
Self-Distillation Methods	Refine models using predictions from their intermediate layers as supervisory signals, improving performance without external teacher models.	[163,164]
RL for Optimization	Explores RL-based strategies to automate pruning and quantization, achieving optimal compression and performance trade-offs.	[165,166]
Cooperative Compression	Discusses collaborative approaches like joint optimization of pruning, quantization, and distillation for maximum resource utilization and scalability.	[167]

4. Advanced Deep Learning Models

The rapid progression of DL in image processing has been marked by the continuous development of advanced models that address the limitations of earlier architectures while introducing novel capabilities. These models represent the cutting edge of DL, incorporating sophisticated mechanisms that enable them to tackle increasingly complex and varied image processing tasks with greater accuracy, efficiency, and adaptability.

4.1. Deep Residual Networks and Beyond

ResNets are a seminal advancement in DL architecture, primarily addressing the degradation problem that arises when training deep NNs. As networks deepen, they often struggle with vanishing and exploding gradients, leading to deteriorated performance. ResNets tackle this issue by introducing skip connections, which allow the network to learn residual functions instead of directly mapping inputs to outputs. This approach enables the training of networks with hundreds of layers, significantly improving performance on complex image-processing tasks like classification and detection [168–172].

ResNets have inspired further innovations, such as ResNeXt and DenseNet, which expand on the concept of residual learning [173]. ResNeXt utilizes a split–transform–merge strategy to aggregate transformations, enhancing the model’s ability to capture diverse features, while DenseNet connects each layer to every other layer in a feed-forward manner, promoting feature reuse and improving efficiency. However, the depth and complexity of these models also introduce challenges [174,175].

4.2. Attention Mechanisms and Transformers

Attention mechanisms have revolutionized DL by enabling models to focus dynamically on the most relevant parts of the input data. Initially developed in the context of natural language processing, attention mechanisms have been adapted for image processing, where they enhance the ability to model complex spatial dependencies. Unlike traditional convolutional networks that apply fixed filters uniformly across the entire image, attention mechanisms weigh different regions according to their importance for the task at hand, significantly improving performance in tasks such as classification, object detection, and segmentation [176–180].

The introduction of ViTs marks a significant leap forward in leveraging attention mechanisms for image processing. ViTs treat images as sequences of patches, using self-attention to model long-range dependencies across the entire image [181]. This capability allows transformers to capture global context in a more flexible and scalable manner than traditional CNNs. However, transformers are computationally expensive, particularly as input sizes increase, posing challenges for real-time applications and deployments on edge devices. Future advancements may focus on hybrid models that combine the strengths of CNNs and transformers, offering a balance between local feature extraction and global context modeling [182–185].

4.3. Generative Models and Adversarial Networks

Generative models, especially GANs, have introduced new dimensions to DL in image processing. GANs consist of two NNs—the generator and the discriminator—engaged in a dynamic adversarial process where the generator creates synthetic images, and the discriminator attempts to distinguish between real and generated images. This interplay enables the generator to produce increasingly realistic images, making GANs highly effective for tasks such as image synthesis, style transfer, and super-resolution [186–189].

GANs have several limitations [190] that can hinder their application and performance. One major challenge is training instability, as the adversarial nature between the generator and discriminator often leads to convergence issues, making it difficult to achieve a balance between the two networks. Another common problem is mode collapse, where the generator produces limited or repetitive outputs instead of capturing the full diversity of the target data distribution. Techniques such as WGANs [191], which optimize a more stable loss function, and CGANs [192], which allow for controlled image generation based on auxiliary information, have been developed to address these issues. Despite the

advancements, GANs require large and diverse datasets to perform effectively, limiting their applicability in data-scarce environments [193,194].

Also, GANs are highly sensitive to hyperparameter settings, requiring meticulous tuning of learning rates, batch sizes, and other factors to ensure optimal performance. Furthermore, GANs typically demand substantial computational resources, especially for tasks involving high-resolution images or complex data distributions. Moreover, GANs are prone to overfitting, where they memorize the training data instead of generalizing to new inputs, which can limit their effectiveness in real-world applications. These limitations highlight the need for careful design, training strategies, and evaluation methods when working with GANs. Lastly, evaluating GANs is also a significant challenge since there is no universally accepted metric to comprehensively assess their output quality and diversity [195].

4.4. Hybrid and Multi-Modal Models

Hybrid and multi-modal models represent a significant advancement in image processing, combining the strengths of different NN architectures or integrating diverse data modalities to enhance performance. For instance, hybrid models that integrate CNNs with transformers capture both spatial features in images and global dependencies, making them particularly effective for complex tasks such as video analysis or visual question answering [196–201].

In multi-modal settings, combining visual data with textual, auditory, or sensory inputs can significantly improve a model's understanding and decision-making capabilities [202]. Effective fusion strategies, such as cross-modal transformers [203] and co-attentive networks [204], are critical for ensuring that combined data contribute meaningfully to the model's performance. These advanced models are particularly impactful in applications such as medical diagnostics, where integrating imaging data with clinical records can lead to more accurate diagnoses, or in autonomous driving, where combining visual, LiDAR, and radar data enhances perception and decision-making processes [205,206].

Table 11 provides a comprehensive classification of the references discussed in this section, noting key advanced DL models across various categories and the contributions of each model type, offering insights into their unique functionalities and the specific challenges they address within the realm of image processing.

Table 11. Summary of advanced DL models.

Model	Description	References
ResNets	Introduced skip connections to address vanishing gradient problems, enabling the training of very deep networks and improving performance in tasks like classification and detection.	[168–172]
ResNeXt	Utilized a split–transform–merge strategy to aggregate transformations, enhancing feature diversity and capture efficiency.	[173]
DenseNet	Introduced densely connected layers to promote feature reuse, reduce the number of parameters, and improve computational efficiency and accuracy.	[174,175]
Attention Mechanisms	Focus dynamically on the most relevant input regions, enhancing spatial dependency modeling for tasks like classification, segmentation, and detection.	[176–180]

Table 11. Cont.

Model	Description	References
Self-Attention	Captures long-range dependencies by relating all elements within a sequence, boosting spatial and temporal understanding for image processing tasks.	[181]
ViTs	Treat images as sequences of patches and use self-attention mechanisms for global dependency modeling and scalability with large datasets.	[182–184]
Hybrid Architectures	Combine CNNs for local feature extraction with transformers for global context modeling, achieving improved performance for complex image tasks.	[185]
GANs	Involves a generator and a discriminator in an adversarial process to synthesize realistic synthetic images, with applications in style transfer and super-resolution.	[186–190]
WGANs	Introduces a stable loss function to mitigate mode collapse and training instability issues in GANs.	[191]
CGANs	Enable controlled image generation using auxiliary information, improving specific task performance like cross-domain synthesis.	[192]
GANs challenges	Limitations such as training instability, mode collapse, and computational resource demands, emphasizing the need for careful design.	[193–195]
Hybrid Models	Combine CNNs with transformers to leverage local feature extraction and global dependency modeling for tasks like video analysis and visual question answering.	[196–201]
Multi-Modal Models	Integrate visual, textual, auditory, or sensory data to enhance understanding and decision-making in tasks like medical diagnostics and autonomous driving.	[202,205,206]
Cross-Modal Transformers	Employ transformers for fusing multiple modalities, such as infrared and visible image fusion, enhancing model adaptability and performance across data modalities.	[203]
Co-Attention Fusion Networks	Focus on aligning multi-modal data streams for specific tasks like multimodal skin cancer diagnosis, improving feature integration and decision accuracy.	[204]

5. Evaluation Metrics for Image Processing Models

Evaluating the performance of DL models in image processing requires a set of well-defined and sophisticated metrics that accurately reflect the quality and efficacy of the models across different tasks. Given the diversity of tasks within image processing—ranging from classification and detection to segmentation and generation—each type of task necessitates specific metrics tailored to its unique requirements. In this section, we delve into the most critical evaluation metrics, providing rigorous mathematical definitions and equations, along with a discussion of their significance and application [207–212].

Accuracy is one of the most fundamental metrics for evaluating image classification tasks. It measures the proportion of correctly classified instances out of the total instances in the dataset. Mathematically, accuracy is defined as follows:

$$\begin{aligned}\text{Accuracy} &= \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} \\ &= \frac{TP + TN}{TP + TN + FP + FN}\end{aligned}\quad (1)$$

where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives, and FN is the number of false negatives. Accuracy is most effective when the dataset is balanced; however, in cases of class imbalance, it may not provide a true reflection of model performance.

Precision and recall are crucial metrics for evaluating image processing tasks such as object detection and segmentation, where class imbalance is common. Precision measures the accuracy of positive predictions and is defined as follows:

$$\text{Precision} = \frac{TP}{TP + FP}. \quad (2)$$

Mean average precision (mAP) is widely used in object detection tasks to evaluate the precision–recall trade-off across different recall thresholds. mAP is calculated by averaging the average precision (AP) across all classes. The AP for each class is computed as follows:

$$\text{AP} = \int_0^1 p(r) dr \quad (3)$$

where $p(r)$ is the precision as a function of recall r . Then, mAP is given by

$$\text{mAP} = \frac{1}{N} \sum_{i=1}^N \text{AP}_i, \quad (4)$$

where N is the number of classes. mAP provides a comprehensive evaluation by considering both precision and recall across various thresholds.

Recall, also known as sensitivity or true positive rate, measures the ability of the model to correctly identify all positive instances and is defined as follows:

$$\text{Recall} = \frac{TP}{TP + FN}. \quad (5)$$

Both precision and recall are critical in contexts where the cost of false positives or false negatives is high.

The F1-Score is the harmonic mean of precision and recall, providing a single metric that balances the trade-off between the two. It is particularly useful when the distribution of classes is uneven and a balance between precision and recall is desired. The F1-Score ranges from 0 to 1, with 1 indicating perfect precision and recall, and is defined as follows:

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2TP}{2TP + FP + FN}. \quad (6)$$

Intersection over Union (IoU) is a critical metric for evaluating object detection and segmentation tasks. Mathematically, IoU is defined as follows:

$$\begin{aligned}\text{IoU} &= \frac{\text{Area of Overlap}}{\text{Area of Union}} \\ &= \frac{|A \cap B|}{|A \cup B|},\end{aligned}\quad (7)$$

where the "Area of Overlap" represents the region where the predicted result and the ground truth agree, while the "Area of Union" captures the total area covered by both the predicted result and the ground truth. More specifically, A is the predicted bounding box or segmentation mask, and B is the ground truth bounding box or segmentation mask. An alternative definition of IoU is

$$\text{IoU} = \frac{TP}{TP + FP + FN}, \quad (8)$$

where TP denotes the region correctly predicted as part of the object, FP captures the region predicted as part of the object but is not part of the ground truth, and FN represents the region that belongs to the object in the ground truth but is not captured by the prediction. IoU measures how well the predicted region overlaps with the actual ground truth. IoU values range from 0 to 1, with higher values indicating better model performance. An IoU threshold (e.g., 0.5) is often used to determine whether a detection is considered a true positive. A perfect IoU (IoU = 1) score means that the predicted region perfectly matches the ground truth (no false positives or false negatives). Lower IoU scores (IoU < 1) indicate that there is either over-prediction (false positives) or under-prediction (false negatives).

The Jaccard Index, also known as the Jaccard Similarity Coefficient, is another metric used for segmentation tasks. It is often referred to separately in the context of binary segmentation. The Jaccard Index, like IoU, provides a measure of similarity between the predicted and ground truth masks, with values ranging from 0 (no overlap) to 1 (perfect overlap). The IoU and Jaccard Index are mathematically identical metrics used to measure the similarity between predicted and ground truth regions, particularly in segmentation tasks. The Jaccard Index originates from set theory as a general measure of set similarity, while IoU is a term more commonly used in computer vision, specifically for tasks like object detection and segmentation. In binary segmentation, the terms are often used interchangeably, but the Jaccard Index is sometimes highlighted separately to emphasize its historical roots and application in simple binary tasks. In contrast, IoU is more widely associated with multi-class scenarios, where mean IoU (average IoU across all classes) is often discussed, and performance thresholds, such as in object detection. Despite these contextual differences, they provide identical numerical evaluations of overlap quality.

The Dice Coefficient is a metric similar to IoU but is specifically tailored for evaluating segmentation tasks (emphasizing the overlap more strongly). It measures the overlap between two samples and is particularly useful in scenarios where the target object occupies a small area in the image. While IoU is widely used in general computer vision tasks, the Dice Coefficient is often favoured in applications like medical imaging due to its sensitivity to smaller regions. Both provide useful insights into the quality of segmentation models. The Dice Coefficient is defined as follows:

$$\text{Dice Coefficient} = \frac{2 \times |A \cap B|}{|A| + |B|}, \quad (9)$$

where A is the predicted segmentation mask, and B is the ground truth segmentation mask. The Dice Coefficient quantifies the degree of similarity by comparing the size of the overlap (True Positives) to the combined size of the predicted and actual regions. The formula is given by

$$\text{Dice Coefficient} = \frac{2TP}{2TP + FP + FN}, \quad (10)$$

where TP are pixels correctly identified as part of the target object, FP the pixels incorrectly identified as belonging to the object, and FN are pixels that belong to the object but were missed in the prediction. A Dice Coefficient of 1 indicates perfect agreement between the prediction and ground truth, while 0 represents no overlap. This metric is particularly

effective in segmentation tasks where precise boundary matching is crucial, such as in medical imaging or autonomous systems.

Pixel accuracy is a straightforward metric used in segmentation tasks. It measures the proportion of correctly classified pixels over the total number of pixels in the image:

$$\text{Pixel Accuracy} = \frac{\sum_{i=1}^N \mathbb{I}(\hat{y}_i = y_i)}{N}, \quad (11)$$

where N is the total number of pixels, \hat{y}_i is the predicted label for pixel i , and y_i is the ground truth label for pixel i . While easy to compute, pixel accuracy may not be sufficient in cases where the classes are imbalanced, as it could overestimate the performance by ignoring small but critical regions.

The Structural Similarity Index (SSIM) is a perceptual metric that quantifies the similarity between two images. Unlike traditional metrics that measure absolute errors, SSIM takes into account changes in structural information, luminance, and contrast. It is defined as follows:

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \quad (12)$$

where x and y are the two images being compared, μ_x and μ_y are the mean intensities of x and y , σ_x^2 and σ_y^2 are the variances of x and y , σ_{xy} is the covariance of x and y , and C_1 and C_2 are constants to avoid division by zero. SSIM values range from -1 to 1 , with higher values indicating greater structural similarity.

The Fréchet Inception Distance (FID) is a metric used to evaluate the quality of images generated by models like GANs. FID compares the distribution of generated images with that of real images using the features extracted from a pre-trained network, typically the Inception model. It is defined as follows:

$$\text{FID} = \|\mu_r - \mu_g\|^2 + \text{Tr}(\Sigma_r + \Sigma_g - 2(\Sigma_r \Sigma_g)^{1/2}), \quad (13)$$

where μ_r and μ_g are the mean feature vectors for the real and generated images, respectively, and Σ_r and Σ_g are the covariance matrices for the real and generated images, respectively. Lower FID values indicate that the generated images are more similar to the real images, with values closer to zero being ideal.

The peak signal-to-noise ratio (PSNR) is a metric used to measure the quality of reconstruction in tasks like image super-resolution and compression. It compares the maximum possible signal to the noise affecting the fidelity of its representation, calculated as follows:

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{\text{MAX}_I^2}{\text{MSE}} \right), \quad (14)$$

where MAX_I is the maximum possible pixel value of the image (e.g., 255 for an 8-bit image), and MSE is the mean squared error between the original and reconstructed images. Higher PSNR values indicate better reconstruction quality.

Normalized Cross-Correlation (NCC) is used in template matching and registration tasks, measuring the similarity between two images. It is defined as follows:

$$\text{NCC} = \frac{\sum_i (I_i - \bar{I})(T_i - \bar{T})}{\sqrt{\sum_i (I_i - \bar{I})^2 \sum_i (T_i - \bar{T})^2}}, \quad (15)$$

where I_i and T_i are the intensity values of the image and template, respectively, and \bar{I} and \bar{T} are the mean intensities of the image and template. NCC values range from -1 to 1 , where 1 indicates perfect correlation.

Cohen's Kappa is a statistical measure of inter-rater agreement or reliability, often used in classification tasks to assess the agreement between predicted and true classifications beyond chance. It is defined as follows:

$$\kappa = \frac{p_o - p_e}{1 - p_e}, \quad (16)$$

where p_o is the observed agreement, and p_e is the expected agreement by chance. Cohen's Kappa ranges from -1 to 1 , with 1 indicating perfect agreement and values less than 0 indicating agreement worse than chance.

The Receiver Operating Characteristic (ROC) curve is a graphical plot that illustrates the diagnostic ability of a binary classifier system by varying its discrimination threshold. The Area Under the ROC Curve (AUC) provides a single scalar value to summarize the overall performance of the classifier:

$$\text{AUC} = \int_0^1 \text{TPR}(\text{FPR}) d(\text{FPR}), \quad (17)$$

where TPR is the true positive rate, and FPR is the false positive rate. AUC values range from 0 to 1 , with values closer to 1 indicating better model performance.

Logarithmic Loss, or Log Loss, measures the performance of a classification model where the prediction is a probability value between 0 and 1 . The log loss increases as the predicted probability diverges from the actual label:

$$\text{Log Loss} = -\frac{1}{N} \sum_{i=1}^N [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)], \quad (18)$$

where N is the number of instances, y_i is the actual label (0 or 1), and p_i is the predicted probability of the instance being in class 1 . Lower Log Loss values indicate better performance.

Lastly, Balanced Accuracy and Matthews Correlation Coefficient (MCC) are advanced metrics used in cases of imbalanced datasets. Balanced accuracy is the average of recall obtained in each class:

$$\text{Balanced Accuracy} = \frac{1}{2} \left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP} \right), \quad (19)$$

MCC provides a comprehensive metric that considers all four quadrants of the confusion matrix:

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}. \quad (20)$$

MCC ranges from -1 to 1 , where 1 indicates perfect prediction, 0 indicates no better than random prediction, and -1 indicates total disagreement between predictions and actual outcomes.

The choice of evaluation metrics is crucial for accurately assessing the performance of DL models in image processing. Each metric provides unique insights into different aspects of model performance, from accuracy and precision to structural similarity and generation quality.

Table 12 provides an overview of the evaluation metrics discussed in this section, categorizing them based on their application to different tasks in image processing. These metrics play a crucial role in assessing the performance and reliability of DL models in diverse scenarios. For instance, classification metrics such as accuracy, precision, recall, and F1-Score are widely used in tasks like object recognition and disease classification, where the balance between false positives and false negatives is critical. In segmentation tasks, metrics like IoU and Dice Coefficient are essential for evaluating the overlap between

predicted and ground truth masks, particularly in medical imaging applications, such as tumor detection, where precise boundaries are crucial. For image quality assessment, metrics such as SSIM and PSNR are ideal for evaluating reconstruction tasks, such as super-resolution or denoising, where perceptual similarity matters. In object detection, metrics like mAP are commonly used to evaluate how well models identify and localize objects in scenes, as seen in autonomous driving systems. Lastly, advanced evaluation metrics like FID are indispensable for assessing the realism of generated images in applications involving generative models. By categorizing these metrics and providing practical guidance, Table 12 serves as a source for selecting the most appropriate evaluation metrics for specific image processing tasks.

Table 12. Grouped evaluation metrics.

Category	Metric
Classification	Accuracy
	Precision
	Recall (Sensitivity)
	F1-Score
	AUC-ROC
	Log Loss
Segmentation and Detection	IoU
	Dice Coefficient
	Jaccard Index
	Pixel Accuracy
Image Quality	SSIM
	PSNR
	NCC
Object Detection Metrics	mAP
Agreement	Cohen's Kappa
	MCC
Advanced Evaluation	Balanced Accuracy
	FID

6. Applications of Deep Learning in Image Processing

DL has profoundly impacted a wide array of domains through its ability to process and interpret complex visual data. Its applications span numerous fields, from healthcare to autonomous systems, each benefiting from the unique capabilities of DL models. This section explores some of the most significant and transformative applications, demonstrated in Table 13, highlighting recent advancements, ethical considerations, and interdisciplinary collaborations.

Table 13. Classification of references related to application fields of DL in image processing.

Topic	References	Description
Medical Imaging	[213–221]	Discusses the revolutionary impact of CNNs on medical diagnostics, such as cancer detection, Alzheimer’s, and diabetic retinopathy. It also highlights how DL models aid in treatment planning and patient monitoring.
Health monitoring	[222–228]	Continuous health monitoring using wearable devices and self-supervised learning and AI, addressing challenges such as bias and disparities in diagnostic accuracy across different demographic groups.
Autonomous systems	[229–235]	Discusses DL applications such as object detection, lane-keeping, and obstacle avoidance in self-driving cars, focusing on real-time decision-making systems like ADASs.
AI and edge computing in autonomous systems	[236–240]	Recent advancements in real-time AI and edge computing enhance autonomous systems’ efficiency.
Remote Sensing and Environmental Monitoring	[241–247]	Discusses the applications of DL in analyzing satellite and aerial imagery, particularly for tracking deforestation, wildlife, damage assessment from natural disasters, and predicting crop yields.
Enhancing Environmental Monitoring	[248–253]	Highlights the integration of DL with remote sensing for environmental decision-making, as well as challenges like the computational cost of processing high-resolution images.
Security and Surveillance	[254–260]	Pertains to the role of DL in real-time video surveillance, facial recognition, and anomaly detection.
Security Surveillance and Ethical Concerns	[261–267]	Discusses the ethical implications of deploying DL in surveillance systems, including concerns around privacy, the potential misuse of technology, and bias.
Art and Cultural Heritage	[268–273]	Describe the applications of DL in restoring damaged artwork, colorizing old photographs, and digitizing cultural artifacts.
AI Collaboration in Cultural Preservation	[274–278]	Focuses on interdisciplinary collaboration between art historians and AI researchers to ensure DL respects cultural integrity and enhances public engagement.
Ethical and Social Considerations	[279–292]	Addresses bias in DL models, particularly in medical imaging and surveillance, as well as the need for fairness, transparency, and accountability in AI systems and privacy-preserving algorithms.
Interdisciplinary Collaboration	[293–304]	Highlights the importance of collaboration between AI researchers and domain experts in healthcare, environmental science, and security for advancing DL applications.

6.1. Medical Imaging

Medical imaging has been one of the most impactful areas for the application of DL in image processing. CNNs have revolutionized diagnostic processes, enabling the detection of diseases such as cancer, Alzheimer’s, and diabetic retinopathy with remarkable accuracy. For instance, DL models have been developed to identify early-stage tumours in mammograms that might be missed by the human eye. These advancements extend beyond diagnosis to treatment planning and monitoring, where models assist in delineating tumours in radiotherapy and predicting patient outcomes [213–221].

Recent innovations, such as self-supervised learning, are further enhancing the field, enabling models to learn from vast amounts of unlabeled medical images, which are often more abundant than labeled data. Additionally, there is a growing integration of AI with wearable technology, facilitating continuous monitoring and early detection of health issues. However, these advancements also bring challenges, particularly in terms of bias in training data that can lead to disparities in diagnostic accuracy across different demographic groups. Addressing these biases, ensuring model interpretability, and adhering to stringent regulatory standards are crucial for the responsible deployment of AI in healthcare [222–228].

6.2. Autonomous Systems

Autonomous vehicles rely heavily on DL for a variety of tasks, including object detection, lane keeping, and obstacle avoidance. The ability of DL models to process real-time video data and make split-second decisions is critical for the safe operation of these vehicles. For example, DL is central to the functioning of advanced driver-assistance systems (ADASs) in vehicles from various companies, where models must accurately detect and respond to pedestrians, other vehicles, and road signs under varying environmental conditions [229–235].

Recent developments in real-time AI and edge computing have further improved the efficiency and reliability of autonomous systems. By processing data closer to the source, edge computing reduces latency and enables faster decision-making, which is crucial in dynamic driving environments. However, significant challenges remain, particularly in ensuring that models generalize well across diverse and unpredictable driving scenarios. Collaboration between AI researchers, automotive engineers, and policymakers is essential to address these challenges and advance the field [236–240].

6.3. Remote Sensing and Environmental Monitoring

Remote sensing and environmental monitoring have also greatly benefited from DL, particularly in the analysis of satellite and aerial imagery. DL models are used to monitor deforestation, track wildlife, assess damage from natural disasters, and predict crop yields. For instance, during disaster response, these models can quickly analyze satellite images to assess the extent of damage and identify areas in need of immediate aid [241–247].

The integration of DL with remote sensing has enabled more accurate and timely decision-making, which is crucial in managing environmental challenges and responding to natural disasters. Moreover, the advent of self-supervised and semi-supervised learning techniques is allowing models to better handle the vast amounts of unlabeled data typical in this field. However, the computational cost of processing high-resolution satellite images remains a challenge, and there is ongoing research into making these models more efficient and scalable [248–253].

6.4. Security and Surveillance

Security and surveillance is another domain where DL is making significant strides. From facial recognition systems to automated threat detection in public spaces, DL models are increasingly being deployed to enhance security. These systems can analyze vast amounts of video data in real time, identifying potential threats and reducing the burden on human operators [254–260].

However, the deployment of DL in surveillance raises serious ethical concerns, particularly regarding privacy and civil liberties. The potential for misuse of facial recognition technology by governments or corporations, as well as the risk of bias in these systems, which could lead to discriminatory practices, are significant challenges that need to be addressed. Research is focused on developing privacy-preserving algorithms and ensuring that these technologies are used in a manner that respects individual rights and freedoms [261–267].

6.5. Art and Cultural Heritage

Art and cultural heritage preservation is a more unconventional but equally important application of DL. Models are being used to restore damaged artworks, colorize black-and-white photographs, and even generate new art in the style of famous artists. DL is also helping to digitize and analyze vast collections of cultural artifacts, making them more accessible to the public and preserving them for future generations [268–273].

In this domain, the focus is not only on technological advancement but also on interdisciplinary collaboration. Art historians, conservators, and AI researchers are working together to ensure that the application of DL respects the integrity and cultural significance of the artifacts. Additionally, there is a growing interest in using AI to enhance the public's engagement with art and culture through interactive and immersive experiences [274–278].

6.6. Ethical and Social Considerations

Across all these applications, ethical and social considerations are paramount. The deployment of DL technologies raises important questions about privacy, bias, and fairness. For example, in the context of surveillance and security, there is a significant risk that these technologies could infringe on individual privacy or be used in ways that exacerbate social inequalities. Similarly, in medical imaging, bias in training data can lead to disparities in diagnosis and treatment outcomes across different demographic groups [279–285].

To address these concerns, it is crucial to develop frameworks and standards that ensure the responsible use of AI. This includes implementing privacy-preserving techniques, designing algorithms that are fair and unbiased, and ensuring transparency and accountability in AI systems. The ethical deployment of DL technologies requires a careful balance between innovation and the protection of fundamental human rights [286–292].

6.7. Interdisciplinary Collaboration

The successful application of DL in image processing often requires interdisciplinary collaboration. In many of the domains discussed, the most impactful advancements have come from partnerships between experts in computer science, domain-specific fields (such as medicine or environmental science), ethics, and law. For instance, in healthcare, the collaboration between AI researchers and clinicians is crucial for developing models that are not only accurate but also clinically relevant and ethically sound [293–298].

Interdisciplinary collaboration ensures that the application of DL is informed by a deep understanding of the context in which it is deployed, leading to more effective and responsible AI solutions. By bringing together diverse perspectives and expertise, these collaborations can help to address complex challenges and maximize the benefits of DL across various domains [299–304].

7. Challenges and Future Directions

As DL continues to revolutionize the field of image processing, it faces several significant challenges that must be addressed to ensure the development of robust, scalable, and ethically sound models. These challenges also open up avenues for future research and innovation, as the field evolves to meet the growing demands of various applications.

7.1. Challenges

One of the foremost challenges in DL for image processing is data scarcity, particularly in specialized domains such as medical imaging, autonomous vehicles, and satellite imagery. In these areas, obtaining large, annotated datasets is not only difficult but also costly, requiring expert knowledge for accurate labeling. This scarcity hinders the training of DL models, which typically require vast amounts of data to achieve high performance. Although techniques like data augmentation and synthetic data generation have been employed to mitigate this issue, they often fall short of providing the diversity and realism needed for truly effective model training [305–310].

Another critical challenge is the computational complexity of DL models. As models grow in size and complexity, they demand significant computational resources for both training and inference. This becomes a major hurdle when deploying models on edge devices or in real-time applications where computational power is limited. Furthermore, the

energy consumption of large-scale models is increasingly becoming a concern, particularly in the context of sustainable AI practices [311–316].

Interpretability remains a significant barrier to the widespread adoption of DL in critical fields such as healthcare, finance, and law. The “black-box” nature of many DL models means that their decision-making processes are often opaque, making it difficult to trust and validate their outputs. This lack of transparency can lead to resistance from stakeholders and regulatory bodies, who require clear justifications for the decisions made by AI systems. The challenge here is not only to develop more interpretable models but also to balance interpretability with performance, as increasing one often comes at the expense of the other [317–322].

Generalization and robustness are also ongoing challenges in DL. Models that perform exceptionally well on training data often struggle to maintain that performance on unseen data, particularly when there is a shift in the data distribution or when the models are exposed to adversarial examples. Ensuring that models generalize well across different environments and are robust to variations and attacks is critical for their reliable deployment in real-world applications [323–328].

Lastly, the ethical implications of deploying DL models in image processing applications cannot be overlooked. Bias in training data can lead to models that reinforce or exacerbate existing societal inequalities, particularly in applications like facial recognition and predictive policing. Privacy concerns arise when AI is used in surveillance or other contexts where sensitive personal information is processed. Addressing these ethical challenges requires a concerted effort to develop fair, transparent, and accountable AI systems [329–334].

A detailed overview of the primary challenges discussed in this section is provided in Table 14, encompassing data scarcity, computational complexity, interpretability, generalization, and ethical considerations. These challenges represent critical hurdles in the development and deployment of effective DL models for image processing, as they impact the scalability, reliability, and transparency of these technologies. Finally, the table synthesizes key references, offering a structured foundation to understand the scope and implications of each challenge within this rapidly evolving field.

Table 14. Classification of references related to DL-based challenges in image processing.

Challenges	References
Data scarcity, particularly in medical imaging, autonomous vehicles, and satellite imagery	[305–310]
Computational complexity of DL models and the challenge of deployment on edge devices	[311–316]
Interpretability challenges, especially in healthcare, finance, and law	[317–322]
Generalization and robustness challenges in DL models	[323–328]
Ethical implications, including bias and privacy concerns in AI systems	[329–334]

7.2. Future Directions

To address these challenges, several promising directions for future research and development have emerged. One of the most significant is the advancement of self-supervised learning techniques. By leveraging vast amounts of unlabeled data, self-supervised learning can help alleviate the issue of data scarcity, allowing models to learn useful representations without the need for extensive labeled datasets. This approach not only reduces the reliance on labeled data but also enhances the model’s ability to generalize across different tasks and domains [335–340].

The development of more efficient model architectures is another critical area of focus. Innovations such as neural architecture search (NAS), pruning, quantization, and distillation are driving the creation of models that are both powerful and computationally efficient. These techniques enable the deployment of DL models on edge devices and in real-time applications, broadening the accessibility and applicability of AI. Additionally, exploring new hardware paradigms, such as neuromorphic and quantum computing, could further revolutionize how DL models are designed and deployed [341–346].

Explainable AI (XAI) is becoming increasingly important as we seek to build trust in AI systems. Research into methods that can make DL models more interpretable without sacrificing performance is gaining momentum. Techniques such as attention mechanisms, feature attribution methods, and interpretable model architectures are crucial for creating AI systems that are transparent and trustworthy. Additionally, developing standards for AI explainability and integrating them into regulatory frameworks will be essential for the broader adoption of AI in sensitive fields [347–353].

Another exciting direction is the integration of emerging technologies with DL. Quantum computing, for example, holds the potential to exponentially accelerate certain computations, making it possible to train and deploy much larger and more complex models. Edge computing, which brings computation closer to the data source, could revolutionize real-time image-processing tasks by reducing latency and improving privacy. The convergence of these technologies with DL could lead to groundbreaking innovations in areas such as autonomous vehicles, smart cities, and personalized medicine [354–359].

Finally, as AI becomes increasingly pervasive, there is a growing need to develop new evaluation metrics that go beyond traditional accuracy and performance measures. These metrics should capture aspects such as robustness, fairness, and ethical considerations, ensuring that models are not only technically sound but also socially responsible. The development of such metrics, along with frameworks for continuous monitoring and auditing of AI systems, will be crucial for ensuring that AI technologies are aligned with societal values [360–365].

In summary, Table 15 outlines the promising future directions explored in this section. This includes advancements in self-supervised learning, efficient model architectures, explainable AI, integration with emerging technologies, and the development of new evaluation metrics. While the challenges facing DL in image processing are significant, they also present opportunities for innovation. By advancing research in the aforementioned directions, the field can continue to evolve, addressing the limitations of current approaches and opening up new possibilities for the future [366–372].

Table 15. Classification of references related to DL-based future directions in image processing.

Future Directions	References
Self-supervised learning and data scarcity	[335–340]
Efficient model architectures, NAS, pruning, quantization, distillation, neuromorphic and quantum computing	[341–346]
XAI and methods for making DL models interpretable	[347–353]
Integration of emerging technologies with DL, quantum computing, and edge computing	[354–359]
Development of new evaluation metrics, fairness, and ethical considerations	[360–365]
Opportunities for innovation in self-supervised learning, efficient model architectures, explainable AI, and emerging technologies	[366–372]

8. Conclusions

DL has fundamentally transformed the landscape of image processing, driving unprecedented advancements across various domains. This survey has provided a comprehensive examination of the key models, techniques, and evaluation metrics that have propelled DL to the forefront of image processing research and application. By tracing the evolution of DL architectures from their inception to the latest state-of-the-art models, we have highlighted the critical innovations that have enabled these models to achieve remarkable success in handling complex visual data.

This survey has underscored the importance of advanced techniques that enhance model performance, such as automated feature extraction, transfer learning, and attention mechanisms. These techniques have not only improved the accuracy and generalization capabilities of DL models but have also expanded their applicability to a wide range of image processing tasks, from basic image recognition to sophisticated tasks like semantic segmentation and image generation.

Furthermore, we have explored the metrics used to evaluate these models, emphasizing the need for rigorous and context-specific assessment to ensure that DL models meet the high standards required for real-world deployment. The discussion on evaluation metrics highlights the nuanced understanding needed to interpret model performance accurately, particularly in diverse and challenging application scenarios.

This survey has also identified the persistent challenges that continue to hinder the full potential of DL in image processing. Issues such as data scarcity, high computational costs, and the black-box nature of DL models present significant obstacles that must be addressed to further advance the field. These challenges underscore the importance of ongoing research into more efficient, interpretable, and accessible DL methodologies.

Looking forward, the integration of DL with emerging technologies such as edge computing, quantum computing, and self-supervised learning offers exciting possibilities for the future of image processing. These advancements have the potential to overcome current limitations, enabling more efficient, scalable, and interpretable models that can be deployed across a wider array of applications, even in resource-constrained environments.

While this survey offers a comprehensive overview of DL techniques and models in image processing, it has several limitations that should be acknowledged. This study primarily focuses on established and recent advancements, potentially under-representing the latest breakthroughs and emerging technologies, such as quantum computing and neuromorphic architectures. Additionally, this survey does not provide in-depth comparative analyses between models under consistent evaluation metrics, limiting practical insights. Interdisciplinary considerations and the role of collaboration in addressing real-world challenges are briefly discussed. Furthermore, while ethical and social implications, such as biases and privacy concerns, are mentioned, they are not explored in depth. These limitations highlight areas for further research, including a more detailed exploration of emerging trends, domain-specific applications, and ethical challenges in deploying DL models.

In summary, this survey not only provides a synthesis of the current state of DL in image processing but also offers a forward-looking perspective on the future directions of the field. By consolidating the vast and diverse body of research into a cohesive overview, this survey serves as a valuable resource for both researchers and practitioners. It lays the groundwork for future innovations, guiding the continued evolution of DL as a transformative force in image processing. The insights presented here aim to inspire further exploration and development, ensuring that DL remains at the cutting edge of image processing technology.

Author Contributions: E.D. and M.T. conceived of the idea, designed and performed the experiments, analyzed the results, drafted the initial manuscript, and revised the final manuscript. All authors have read and agreed to the published version of the manuscript

Funding: This research received no external funding

Conflicts of Interest: The authors declare no conflicts of interest.

List of Abbreviations

The following abbreviations are used in this manuscript:

Acronym	Meaning
AI	Artificial Intelligence
DL	Deep Learning
ML	Machine Learning
GPUs	Graphics Processing Units
CNN	Convolutional Neural Network
ResNet	Residual Network
DesNet	Densely Connected Convolutional Network
FCN	Fully Convolutional Network
R-CNN	Region-based Convolutional Neural Network
YOLO	You Only Look Once
NN	Neural Network
ConvNext	Next Generation of Convolutional Networks
ViT	Vision Transformer
GAN	Generative Adversarial Network
CGAN	Conditional GAN
WGAN	Wasserstein GAN
FGSM	Fast Gradient Sign Method
PGD	Projected Gradient Descent
MAT	Model-based Adversarial Training
UPGD	Universal Projected Gradient Descent
HTPL	Hierarchical Transfer Progressive Learning
RL	Reinforcement Learning
SimCLR	Simple Framework for Contrastive Learning of Visual
DDC	Deep Domain Confusion
DICA	Domain-Invariant Component Analysis
DANN	Domain-Adversarial NN
MANN	Memory-Augmented NN
SNAIL	Neural Attentive Meta-Learner
MAML	Model-Agnostic Meta-Learning
LSTM	Long Short-Term Memory
SGD	Stochastic Gradient Descent
CLIP	Contrastive Language-Image Pretraining
PRISM	Promptable and Robust Interactive Segmentation Model
MoCo	Momentum Contrast
NAS	Neural Architecture Search
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
TPR	True Positive Rate
FPR	False Positive Rate
IoU	Intersection over Union
AP	Average Precision

mAP	Mean AP
SSIM	Structural Similarity Index
FID	Fréchet Inception Distance
PSNR	Peak Signal-to-Noise Ratio
NCC	Normalized Cross-Correlation
ROC	Receiver Operating Characteristic
AUC	Area Under the ROC Curve
MCC	Matthews Correlation Coefficient
ADAS	Advanced Driver-Assistance System
XAI	Explainable AI

References

1. Monga, V.; Li, Y.; Eldar, Y.C. Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing. *IEEE Signal Process. Mag.* **2021**, *38*, 18–44. [\[CrossRef\]](#)
2. Banan, A.; Nasiri, A.; Taheri-Garavand, A. Deep learning-based appearance features extraction for automated carp species identification. *Aquac. Eng.* **2020**, *89*, 102053. [\[CrossRef\]](#)
3. Wang, P.; Fan, E.; Wang, P. Comparative analysis of image classification algorithms based on traditional machine learning and deep learning. *Pattern Recognit. Lett.* **2021**, *141*, 61–67. [\[CrossRef\]](#)
4. Li, L.; Zhou, T.; Wang, W.; Li, J.; Yang, Y. Deep hierarchical semantic segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, USA, 18–22 June; pp. 1246–1257.
5. Li, X.; Wang, T.; Cui, H.; Zhang, G.; Cheng, Q.; Dong, T.; Jiang, B. SARPointNet: An automated feature learning framework for spaceborne SAR image registration. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 6371–6381. [\[CrossRef\]](#)
6. Alshayegi, M.; Al-Buloushi, J.; Ashkanani, A.; Abed, S. Enhanced brain tumor classification using an optimized multi-layered convolutional neural network architecture. *Multimed. Tools Appl.* **2021**, *80*, 28897–28917. [\[CrossRef\]](#)
7. Duan, R.; Deng, H.; Tian, M.; Deng, Y.; Lin, J. SODA: A large-scale open site object detection dataset for deep learning in construction. *Autom. Constr.* **2022**, *142*, 104499. [\[CrossRef\]](#)
8. Jeon, W.; Ko, G.; Lee, J.; Lee, H.; Ha, D.; Ro, W.W. Deep learning with GPUs. In *Advances in Computers*; Elsevier: Amsterdam, The Netherlands, **2021**, *122*, 167–215.
9. Cai, L.; Gao, J.; Zhao, D. A review of the application of deep learning in medical image classification and segmentation. *Ann. Transl. Med.* **2020**, *8*. [\[CrossRef\]](#)
10. Wang, X.; Zhao, Y.; Pourpanah, F. Recent advances in deep learning. *Int. J. Mach. Learn. Cybern.* **2020**, *11*, 747–750. [\[CrossRef\]](#)
11. Liu, Y.; Pu, H.; Sun, D.W. Efficient extraction of deep image features using convolutional neural network (CNN) for applications in detecting and analysing complex food matrices. *Trends Food Sci. Technol.* **2021**, *113*, 193–204. [\[CrossRef\]](#)
12. Hoeser, T.; Kuenzer, C. Object detection and image segmentation with deep learning on earth observation data: A review-part i: Evolution and recent trends. *Remote Sens.* **2020**, *12*, 1667. [\[CrossRef\]](#)
13. Shin, D.; He, S.; Lee, G.M.; Whinston, A.B.; Cetintas, S.; Lee, K.C. *Enhancing Social Media Analysis with Visual Data Analytics: A Deep Learning Approach*; SSRN: Amsterdam, The Netherlands, 2020.
14. Csurka, G.; Hospedales, T.M.; Salzmann, M.; Tommasi, T. *Visual Domain Adaptation in the Deep Learning Era*; Springer: Cham, Switzerland, 2022. [\[CrossRef\]](#)
15. Lilhore, U.K.; Simaiya, S.; Kaur, A.; Prasad, D.; Khurana, M.; Verma, D.K.; Hassan, A. Impact of deep learning and machine learning in industry 4.0: Impact of deep learning. In *Cyber-Physical, IoT, and Autonomous Systems in Industry 4.0*; CRC Press: Boca Raton, FL, USA, 2021; pp. 179–197.
16. Li, X.; Xiong, H.; Li, X.; Wu, X.; Zhang, X.; Liu, J.; Bian, J.; Dou, D. Interpretable deep learning: Interpretation, interpretability, trustworthiness, and beyond. *Knowl. Inf. Syst.* **2022**, *64*, 3197–3234. [\[CrossRef\]](#)
17. Greenwald, N.F.; Miller, G.; Moen, E.; Kong, A.; Kagel, A.; Dougherty, T.; Fullaway, C.C.; McIntosh, B.J.; Leow, K.X.; Schwartz, M.S.; et al. Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. *Nat. Biotechnol.* **2022**, *40*, 555–565. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Thompson, N.C.; Greenewald, K.; Lee, K.; Manso, G.F. The computational limits of deep learning. *arXiv* **2020**, arXiv:2007.05558.
19. Zhan, Z.H.; Li, J.Y.; Zhang, J. Evolutionary deep learning: A survey. *Neurocomputing* **2022**, *483*, 42–58. [\[CrossRef\]](#)
20. Sarwinda, D.; Paradisa, R.H.; Bustamam, A.; Anggia, P. Deep learning in image classification using residual network (ResNet) variants for detection of colorectal cancer. *Procedia Comput. Sci.* **2021**, *179*, 423–431. [\[CrossRef\]](#)
21. Liang, J. Image classification based on RESNET. In *Proceedings of the Journal of Physics: Conference Series*; IOP Publishing: Bristol, UK, 2020; Volume 1634, p. 012110.

22. Yu, D.; Yang, J.; Zhang, Y.; Yu, S. Additive DenseNet: Dense connections based on simple addition operations. *J. Intell. Fuzzy Syst.* **2021**, *40*, 5015–5025. [\[CrossRef\]](#)
23. Chen, B.; Zhao, T.; Liu, J.; Lin, L. Multipath feature recalibration DenseNet for image classification. *Int. J. Mach. Learn. Cybern.* **2021**, *12*, 651–660. [\[CrossRef\]](#)
24. Liu, M.; Chen, L.; Du, X.; Jin, L.; Shang, M. Activated gradients for deep neural networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *34*, 2156–2168. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Khan, S.D.; Basalamah, S. Multi-branch deep learning framework for land scene classification in satellite imagery. *Remote Sens.* **2023**, *15*, 3408. [\[CrossRef\]](#)
26. Alom, M.Z.; Hasan, M.; Yakopcic, C.; Taha, T.M.; Asari, V.K. Inception recurrent convolutional neural network for object recognition. *Mach. Vis. Appl.* **2021**, *32*, 1–14. [\[CrossRef\]](#)
27. Wang, Z.; Wang, Z.; Zeng, C.; Yu, Y.; Wan, X. High-quality image compressed sensing and reconstruction with multi-scale dilated convolutional neural network. *Circuits Syst. Signal Process.* **2023**, *42*, 1593–1616. [\[CrossRef\]](#)
28. Bergamasco, L.; Bovolo, F.; Bruzzone, L. A dual-branch deep learning architecture for multisensor and multitemporal remote sensing semantic segmentation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 2147–2162. [\[CrossRef\]](#)
29. Ragab, M.G.; Abdulkader, S.J.; Muneer, A.; Alqushaibi, A.; Sumiea, E.H.; Qureshi, R.; Al-Selwi, S.M.; Alhussian, H. A Comprehensive Systematic Review of YOLO for Medical Object Detection (2018 to 2023). *IEEE Access* **2024**, *12*, 57815–57836. [\[CrossRef\]](#)
30. Vijayakumar, A.; Vairavasundaram, S. Yolo-based object detection models: A review and its applications. *Multimed. Tools Appl.* **2024**, *83*, 83535–83574. [\[CrossRef\]](#)
31. Qi, J.; Nguyen, M.; Yan, W.Q. Waste classification from digital images using ConvNeXt. In Proceedings of the 10th Pacific-Rim Symposium on Image and Video Technology, Online, 25–28 November; Springer: Berlin/Heidelberg, Germany, 2022; pp. 1–13.
32. Todi, A.; Narula, N.; Sharma, M.; Gupta, U. ConvNext: A Contemporary Architecture for Convolutional Neural Networks for Image Classification. In Proceedings of the 3rd International Conference on Innovative Sustainable Computational Technologies (CISCT), Dehradun, India, 8–9 September 2023; pp. 1–6.
33. Ramos, L.; Casas, E.; Romero, C.; Rivas-Echeverría, F.; Morocho-Cayamcela, M.E. A study of convnext architectures for enhanced image captioning. *IEEE Access* **2024**, *12*, 13711–13728. [\[CrossRef\]](#)
34. Mou, L.; Hua, Y.; Zhu, X.X. Relation matters: Relational context-aware fully convolutional network for semantic segmentation of high-resolution aerial images. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 7557–7569. [\[CrossRef\]](#)
35. Du, G.; Cao, X.; Liang, J.; Chen, X.; Zhan, Y. Medical Image Segmentation based on U-Net: A Review. *J. Imaging Sci. Technol.* **2020**, *64*, 1. [\[CrossRef\]](#)
36. Li, H.; Wang, W.; Wang, M.; Li, L.; Vimlund, V. A review of deep learning methods for pixel-level crack detection. *J. Traffic Transp. Eng. (Engl. Ed.)* **2022**, *9*, 945–968. [\[CrossRef\]](#)
37. Yang, H.; Huang, C.; Wang, L.; Luo, X. An improved encoder–decoder network for ore image segmentation. *IEEE Sensors J.* **2020**, *21*, 11469–11475. [\[CrossRef\]](#)
38. Lin, K.; Zhao, H.; Lv, J.; Li, C.; Liu, X.; Chen, R.; Zhao, R. Face Detection and Segmentation Based on Improved Mask R-CNN. *Discret. Dyn. Nat. Soc.* **2020**, *2020*, 9242917. [\[CrossRef\]](#)
39. Muhammad, K.; Hussain, T.; Ullah, H.; Del Ser, J.; Rezaei, M.; Kumar, N.; Hijji, M.; Bellavista, P.; de Albuquerque, V.H.C. Vision-based semantic segmentation in scene understanding for autonomous driving: Recent achievements, challenges, and outlooks. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 22694–22715. [\[CrossRef\]](#)
40. Nie, X.; Duan, M.; Ding, H.; Hu, B.; Wong, E.K. Attention mask R-CNN for ship detection and segmentation from remote sensing images. *IEEE Access* **2020**, *8*, 9325–9334. [\[CrossRef\]](#)
41. Li, S.; Zhao, X. Pixel-level detection and measurement of concrete crack using faster region-based convolutional neural network and morphological feature extraction. *Meas. Sci. Technol.* **2021**, *32*, 065010. [\[CrossRef\]](#)
42. Udendhran, R.; Balamurugan, M.; Suresh, A.; Varatharajan, R. Enhancing image processing architecture using deep learning for embedded vision systems. *Microprocess. Microsystems* **2020**, *76*, 103094. [\[CrossRef\]](#)
43. Khan, A.; Rauf, Z.; Khan, A.R.; Rathore, S.; Khan, S.H.; Shah, N.S.; Farooq, U.; Asif, H.; Asif, A.; Zahoor, U.; et al. A recent survey of vision transformers for medical image segmentation. *arXiv* **2023**, arXiv:2312.00634.
44. Liu, Q.; Xu, Z.; Bertasius, G.; Niethammer, M. Simpleclick: Interactive image segmentation with simple vision transformers. In Proceedings of the IEEE/CVF International Conference on Computer Vision. Paris, France, 1–6 October 2023; pp. 22290–22300.
45. Qian, X.; Zhang, C.; Chen, L.; Li, K. Deep learning-based identification of maize leaf diseases is improved by an attention mechanism: Self-attention. *Front. Plant Sci.* **2022**, *13*, 864486. [\[CrossRef\]](#)
46. Azad, R.; Kazerouni, A.; Heidari, M.; Aghdam, E.K.; Molaei, A.; Jia, Y.; Jose, A.; Roy, R.; Merhof, D. Advances in medical image analysis with vision transformers: A comprehensive review. *Med. Image Anal.* **2023**, *91*, 103000. [\[CrossRef\]](#) [\[PubMed\]](#)
47. Hassani, A.; Walton, S.; Shah, N.; Abuduweili, A.; Li, J.; Shi, H. Escaping the big data paradigm with compact transformers. *arXiv* **2021**, arXiv:2104.05704.

48. Zhao, H.; Jia, J.; Koltun, V. Exploring self-attention for image recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 10076–10085.
49. Li, S.; Wu, C.; Xiong, N. Hybrid architecture based on CNN and transformer for strip steel surface defect classification. *Electronics* **2022**, *11*, 1200. [[CrossRef](#)]
50. Fang, J.; Lin, H.; Chen, X.; Zeng, K. A hybrid network of cnn and transformer for lightweight image super-resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 1103–1112.
51. Sun, Q.; Fang, N.; Liu, Z.; Zhao, L.; Wen, Y.; Lin, H. HybridCTrm: Bridging CNN and transformer for multimodal brain image segmentation. *J. Healthc. Eng.* **2021**, *2021*, 7467261. [[CrossRef](#)]
52. Akil, M.; Saouli, R.; Kachouri, R. Fully automatic brain tumor segmentation with deep learning-based selective attention using overlapping patches and multi-class weighted cross-entropy. *Med. Image Anal.* **2020**, *63*, 101692.
53. Kumar, V.R.; Yogamani, S.; Milz, S.; Mäder, P. FisheyeDistanceNet++: Self-supervised fisheye distance estimation with self-attention, robust loss function and camera view generalization. *Electron. Imaging* **2021**, *33*, 1–11.
54. Gong, M.; Chen, S.; Chen, Q.; Zeng, Y.; Zhang, Y. Generative adversarial networks in medical image processing. *Curr. Pharm. Des.* **2021**, *27*, 1856–1868. [[CrossRef](#)] [[PubMed](#)]
55. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks. *Commun. ACM* **2020**, *63*, 139–144. [[CrossRef](#)]
56. Christophe, S.; Mermet, S.; Laurent, M.; Touya, G. Neural map style transfer exploration with GANs. *Int. J. Cartogr.* **2022**, *8*, 18–36. [[CrossRef](#)]
57. Chen, H. Challenges and corresponding solutions of generative adversarial networks (GANs): A survey study. In *Proceedings of the Journal of Physics: Conference Series*; IOP Publishing: Bristol, UK, 2021; Volume 1827, p. 012066.
58. Qin, Z.; Liu, Z.; Zhu, P.; Ling, W. Style transfer in conditional GANs for cross-modality synthesis of brain magnetic resonance images. *Comput. Biol. Med.* **2022**, *148*, 105928. [[CrossRef](#)] [[PubMed](#)]
59. Kim, C.; Park, S.; Hwang, H.J. Local stability of wasserstein GANs with abstract gradient penalty. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *33*, 4527–4537. [[CrossRef](#)] [[PubMed](#)]
60. Zeng, Q.; Ma, X.; Cheng, B.; Zhou, E.; Pang, W. Gans-based data augmentation for citrus disease severity detection using deep learning. *IEEE Access* **2020**, *8*, 172882–172891. [[CrossRef](#)]
61. Balaji, Y.; Chellappa, R.; Feizi, S. Robust optimal transport with applications in generative modeling and domain adaptation. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 12934–12944.
62. Figueira, A.; Vaz, B. Survey on synthetic data generation, evaluation methods and GANs. *Mathematics* **2022**, *10*, 2733. [[CrossRef](#)]
63. Kazemina, S.; Baur, C.; Kuijper, A.; van Ginneken, B.; Navab, N.; Albarqouni, S.; Mukhopadhyay, A. GANs for medical image analysis. *Artif. Intell. Med.* **2020**, *109*, 101938. [[CrossRef](#)]
64. Yamaguchi, S.; Kanai, S.; Eda, T. Effective data augmentation with multi-domain learning gans. In Proceedings of the 34th AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 6566–6574.
65. Croitoru, F.A.; Hondru, V.; Ionescu, R.T.; Shah, M. Diffusion models in vision: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 10850–10869. [[CrossRef](#)] [[PubMed](#)]
66. Cao, H.; Tan, C.; Gao, Z.; Xu, Y.; Chen, G.; Heng, P.A.; Li, S.Z. A survey on generative diffusion models. *IEEE Trans. Knowl. Data Eng.* **2024**, *36*, 2814–2830. [[CrossRef](#)]
67. Iman, M.; Arabnia, H.R.; Rasheed, K. A review of deep transfer learning and recent advancements. *Technologies* **2023**, *11*, 40. [[CrossRef](#)]
68. Matsoukas, C.; Haslum, J.F.; Sorkhei, M.; Söderberg, M.; Smith, K. What makes transfer learning work for medical images: Feature reuse & other factors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 9225–9234.
69. Alzubaidi, L.; Fadhel, M.A.; Al-Shamma, O.; Zhang, J.; Santamaría, J.; Duan, Y.; R. Oleiwi, S. Towards a better understanding of transfer learning for medical imaging: A case study. *Appl. Sci.* **2020**, *10*, 4523. [[CrossRef](#)]
70. Alzubaidi, L.; Al-Amidie, M.; Al-Asadi, A.; Humaidi, A.J.; Al-Shamma, O.; Fadhel, M.A.; Zhang, J.; Santamaría, J.; Duan, Y. Novel transfer learning approach for medical imaging with limited labeled data. *Cancers* **2021**, *13*, 1590. [[CrossRef](#)] [[PubMed](#)]
71. Chen, H.; Wang, Y.; Guo, T.; Xu, C.; Deng, Y.; Liu, Z.; Ma, S.; Xu, C.; Xu, C.; Gao, W. Pre-trained image processing transformer. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 12299–12310.
72. Gupta, J.; Pathak, S.; Kumar, G. Deep learning (CNN) and transfer learning: A review. In *Proceedings of the Journal of Physics: Conference Series*; IOP Publishing: Bristol, UK, 2022; Volume 2273, p. 012029.
73. Kim, H.E.; Cosa-Linan, A.; Santhanam, N.; Jannesari, M.; Maros, M.E.; Ganslandt, T. Transfer learning for medical image classification: A literature review. *BMC Med. Imaging* **2022**, *22*, 69. [[CrossRef](#)] [[PubMed](#)]

74. Wang, Z. Mitigating Negative Transfer for Better Generalization and Efficiency in Transfer Learning. Ph.D. Thesis, Carnegie Mellon University, Pittsburgh, PA, USA, 2022.
75. Agarwal, N.; Sondhi, A.; Chopra, K.; Singh, G. Transfer learning: Survey and classification. *Smart Innov. Commun. Comput. Sci. Proc. ICSICCS 2020* **2021**, *1168*, 145–155.
76. Zhang, W.; Deng, L.; Zhang, L.; Wu, D. A survey on negative transfer. *IEEE/CAA J. Autom. Sin.* **2022**, *10*, 305–329. [[CrossRef](#)]
77. Yang, Y.; Huang, L.K.; Wei, Y. Concept-wise Fine-tuning Matters in Preventing Negative Transfer. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 1–6 October 2023; pp. 18753–18763.
78. Chen, X.; Tao, H.; Zhou, H.; Zhou, P.; Deng, Y. Hierarchical and progressive learning with key point sensitive loss for sonar image classification. *Multimed. Syst.* **2024**, *30*, 1–16. [[CrossRef](#)]
79. Yang, S.; Xiao, W.; Zhang, M.; Guo, S.; Zhao, J.; Shen, F. Image data augmentation for deep learning: A survey. *arXiv* **2022**, arXiv:2204.08610.
80. Maharana, K.; Mondal, S.; Nemade, B. A review: Data pre-processing and data augmentation techniques. *Glob. Transit. Proc.* **2022**, *3*, 91–99. [[CrossRef](#)]
81. Xu, M.; Yoon, S.; Fuentes, A.; Park, D.S. A comprehensive survey of image augmentation techniques for deep learning. *Pattern Recognit.* **2023**, *137*, 109347. [[CrossRef](#)]
82. Rebuffi, S.A.; Goyal, S.; Calian, D.A.; Stimberg, F.; Wiles, O.; Mann, T.A. Data augmentation can improve robustness. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 29935–29948.
83. Li, P.; Li, D.; Li, W.; Gong, S.; Fu, Y.; Hospedales, T.M. A simple feature augmentation for domain generalization. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 8886–8895.
84. Mumuni, A.; Mumuni, F. Data augmentation: A comprehensive survey of modern approaches. *Array* **2022**, *16*, 100258. [[CrossRef](#)]
85. Termritthikun, C.; Jamtsho, Y.; Muneesawang, P. An improved residual network model for image recognition using a combination of snapshot ensembles and the cutout technique. *Multimed. Tools Appl.* **2020**, *79*, 1475–1495. [[CrossRef](#)]
86. Galdran, A.; Carneiro, G.; González Ballester, M.A. Balanced-mixup for highly imbalanced medical image classification. In Proceedings of the 24th International Conference on Medical Image Computing and Computer Assisted Intervention, Strasbourg, France, 27 September–1 October 2021; pp. 323–333.
87. Walawalkar, D.; Shen, Z.; Liu, Z.; Savvides, M. Attentive cutmix: An enhanced data augmentation approach for deep learning based image classification. *arXiv* **2020**, arXiv:2003.13048.
88. Yun, J.P.; Shin, W.C.; Koo, G.; Kim, M.S.; Lee, C.; Lee, S.J. Automated defect inspection system for metal surfaces based on deep learning and data augmentation. *J. Manuf. Syst.* **2020**, *55*, 317–324. [[CrossRef](#)]
89. Tian, K.; Lin, C.; Sun, M.; Zhou, L.; Yan, J.; Ouyang, W. Improving auto-augment via augmentation-wise weight sharing. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 19088–19098.
90. Cubuk, E.D.; Zoph, B.; Shlens, J.; Le, Q.V. Randaugment: Practical automated data augmentation with a reduced search space. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 702–703.
91. Moradi, R.; Berangi, R.; Minaei, B. A survey of regularization strategies for deep models. *Artif. Intell. Rev.* **2020**, *53*, 3947–3986. [[CrossRef](#)]
92. Nandini, G.S.; Kumar, A.S.; Chidananda, K. Dropout technique for image classification based on extreme learning machine. *Glob. Transit. Proc.* **2021**, *2*, 111–116. [[CrossRef](#)]
93. Garbin, C.; Zhu, X.; Marques, O. Dropout vs. batch normalization: An empirical study of their impact to deep learning. *Multimed. Tools Appl.* **2020**, *79*, 12777–12815. [[CrossRef](#)]
94. Wu, L.; Li, J.; Wang, Y.; Meng, Q.; Qin, T.; Chen, W.; Zhang, M.; Liu, T.Y.; et al. R-drop: Regularized dropout for neural networks. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 10890–10905.
95. Andriushchenko, M.; D’Angelo, F.; Varre, A.; Flammarion, N. Why Do We Need Weight Decay in Modern Deep Learning? *arXiv* **2023**, arXiv:2310.04415.
96. Li, X.; Chen, S.; Yang, J. Understanding the disharmony between weight normalization family and weight decay. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 4715–4722.
97. De, S.; Smith, S. Batch normalization biases residual blocks towards the identity function in deep networks. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 19964–19975.
98. Awais, M.; Iqbal, M.T.B.; Bae, S.H. Revisiting internal covariate shift for batch normalization. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *32*, 5082–5092. [[CrossRef](#)] [[PubMed](#)]
99. Zhao, W.; Alwidian, S.; Mahmoud, Q.H. Adversarial training methods for deep learning: A systematic review. *Algorithms* **2022**, *15*, 283. [[CrossRef](#)]
100. Allen-Zhu, Z.; Li, Y. Feature purification: How adversarial training performs robust deep learning. In Proceedings of the 2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS), Denver, CO, USA, 7–10 February 2022; pp. 977–988.

101. Chang, C.L.; Hung, J.L.; Tien, C.W.; Tien, C.W.; Kuo, S.Y. Evaluating robustness of ai models against adversarial attacks. In Proceedings of the 1st ACM Workshop on Security and Privacy on Artificial Intelligence, Taipei, Taiwan, 6 October 2020; pp. 47–54.
102. Silva, S.H.; Najafirad, P. Opportunities and challenges in deep learning adversarial robustness: A survey. *arXiv* **2020**, arXiv:2007.00753.
103. Xie, C.; Tan, M.; Gong, B.; Wang, J.; Yuille, A.L.; Le, Q.V. Adversarial examples improve image recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 819–828.
104. Naqvi, S.M.A.; Shabaz, M.; Khan, M.A.; Hassan, S.I. Adversarial attacks on visual objects using the fast gradient sign method. *J. Grid Comput.* **2023**, *21*, 52. [\[CrossRef\]](#)
105. Lanfredi, R.B.; Schroeder, J.D.; Tasdizen, T. Quantifying the preferential direction of the model gradient in adversarial training with projected gradient descent. *Pattern Recognit.* **2023**, *139*, 109430. [\[CrossRef\]](#) [\[PubMed\]](#)
106. Wong, E.; Rice, L.; Kolter, J.Z. Fast is better than free: Revisiting adversarial training. *arXiv* **2020**, arXiv:2001.03994.
107. Deng, Y.; Karam, L.J. Universal adversarial attack via enhanced projected gradient descent. In Proceedings of the IEEE International Conference on Image Processing (ICIP), Virtual Conference, Abu Dhabi, United Arab Emirates, 25–28 September 2020; pp. 1241–1245.
108. Robey, A.; Hassani, H.; Pappas, G.J. Model-based robust deep learning: Generalizing to natural, out-of-distribution data. *arXiv* **2020**, arXiv:2005.10247.
109. Schmarje, L.; Santarossa, M.; Schröder, S.M.; Koch, R. A survey on semi-, self-and unsupervised learning for image classification. *IEEE Access* **2021**, *9*, 82146–82168. [\[CrossRef\]](#)
110. Yuan, Y.; Wang, C.; Jiang, Z. Proxy-based deep learning framework for spectral-spatial hyperspectral image classification: Efficient and robust. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–15. [\[CrossRef\]](#)
111. Jaiswal, A.; Babu, A.R.; Zadeh, M.Z.; Banerjee, D.; Makedon, F. A survey on contrastive self-supervised learning. *Technologies* **2020**, *9*, 2. [\[CrossRef\]](#)
112. Li, Y.; Chen, J.; Zheng, Y. A multi-task self-supervised learning framework for scopy images. In Proceedings of the 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), Iowa City, IA, USA, 3–7 April, 2020; pp. 2005–2009.
113. Chen, S.; Xue, J.H.; Chang, J.; Zhang, J.; Yang, J.; Tian, Q. SSL++: Improving self-supervised learning by mitigating the proxy task-specificity problem. *IEEE Trans. Image Process.* **2021**, *31*, 1134–1148. [\[CrossRef\]](#)
114. Wang, C.; Wu, Y.; Qian, Y.; Kumatani, K.; Liu, S.; Wei, F.; Zeng, M.; Huang, X. Unispeech: Unified speech representation learning with labeled and unlabeled data. In Proceedings of the International Conference on Machine Learning, Virtual, 18–24 July 2021; pp. 10937–10947.
115. Ericsson, L.; Gouk, H.; Loy, C.C.; Hospedales, T.M. Self-supervised representation learning: Introduction, advances, and challenges. *IEEE Signal Process. Mag.* **2022**, *39*, 42–62. [\[CrossRef\]](#)
116. Chen, X.; Ding, M.; Wang, X.; Xin, Y.; Mo, S.; Wang, Y.; Han, S.; Luo, P.; Zeng, G.; Wang, J. Context autoencoder for self-supervised representation learning. *Int. J. Comput. Vis.* **2024**, *132*, 208–223. [\[CrossRef\]](#)
117. Albelwi, S. Survey on self-supervised learning: Auxiliary pretext tasks and contrastive learning methods in imaging. *Entropy* **2022**, *24*, 551. [\[CrossRef\]](#) [\[PubMed\]](#)
118. Chen, T.; Kornblith, S.; Norouzi, M.; Hinton, G. A simple framework for contrastive learning of visual representations. In Proceedings of the International Conference on Machine Learning, Virtual, 12–18 July 2020; pp. 1597–1607.
119. Ci, Y.; Lin, C.; Bai, L.; Ouyang, W. Fast-MoCo: Boost momentum-based contrastive learning with combinatorial patches. In Proceedings of the European Conference on Computer Vision, Tel Aviv, Israel, 23–27 October 2022; pp. 290–306.
120. He, K.; Fan, H.; Wu, Y.; Xie, S.; Girshick, R. Momentum contrast for unsupervised visual representation learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June, 2020; pp. 9729–9738.
121. Wang, Y.; Zhang, J.; Kan, M.; Shan, S.; Chen, X. Self-supervised equivariant attention mechanism for weakly supervised semantic segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 12275–12284.
122. Diba, A.; Sharma, V.; Safdari, R.; Lotfi, D.; Sarfraz, S.; Stiefelhagen, R.; Van Gool, L. Vi2clr: Video and image for visual contrastive learning of representation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 1502–1512.
123. Allaoui, M.; Kherfi, M.L.; Cheriet, A. Considerably improving clustering algorithms using UMAP dimensionality reduction technique: A comparative study. In Proceedings of the International Conference on Image and Signal Processing, Virtual, 23–25 October 2020; pp. 317–325.
124. Zebari, R.; Abdulazeez, A.; Zeebaree, D.; Zebari, D.; Saeed, J. A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction. *J. Appl. Sci. Technol. Trends* **2020**, *1*, 56–70. [\[CrossRef\]](#)
125. Nalepa, J.; Myller, M.; Imai, Y.; Honda, K.i.; Takeda, T.; Antoniuk, M. Unsupervised segmentation of hyperspectral images using 3-D convolutional autoencoders. *IEEE Geosci. Remote Sens. Lett.* **2020**, *17*, 1948–1952. [\[CrossRef\]](#)
126. Raza, K.; Singh, N.K. A tour of unsupervised deep learning for medical image analysis. *Curr. Med. Imaging* **2021**, *17*, 1059–1077.

127. Rai, S.; Bhatt, J.S.; Patra, S.K. An unsupervised deep learning framework for medical image denoising. *arXiv* **2021**, arXiv:2103.06575.
128. Kim, W.; Kanezaki, A.; Tanaka, M. Unsupervised learning of image segmentation based on differentiable feature clustering. *IEEE Trans. Image Process.* **2020**, *29*, 8055–8068. [[CrossRef](#)]
129. Yoon, J.S.; Oh, K.; Shin, Y.; Mazurowski, M.A.; Suk, H.I. Domain Generalization for Medical Image Analysis: A Review. *Proc. IEEE* **2024**, *112*, 1583–1609. [[CrossRef](#)]
130. Zhou, K.; Liu, Z.; Qiao, Y.; Xiang, T.; Loy, C.C. Domain generalization: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 4396–4415. [[CrossRef](#)] [[PubMed](#)]
131. Zhang, W.; Wang, F.; Jiang, Y.; Xu, Z.; Wu, S.; Zhang, Y. Cross-subject EEG-based emotion recognition with deep domain confusion. In Proceedings of the 12th International Conference on Intelligent Robotics and Applications (ICIRA), Shenyang, China, 8–11 August 2019; pp. 558–570.
132. Wang, F.; Han, Z.; Gong, Y.; Yin, Y. Exploring domain-invariant parameters for source free domain adaptation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 7151–7160.
133. Khoei, A.G.; Yu, Y.; Feldt, R. Domain generalization through meta-learning: A survey. *Artif. Intell. Rev.* **2024**, *57*, 285. [[CrossRef](#)]
134. Sicilia, A.; Zhao, X.; Hwang, S.J. Domain adversarial neural networks for domain generalization: When it works and how to improve. *Mach. Learn.* **2023**, *112*, 2685–2721. [[CrossRef](#)]
135. Liu, Y.; Chen, A.; Shi, H.; Huang, S.; Zheng, W.; Liu, Z.; Zhang, Q.; Yang, X. CT synthesis from MRI using multi-cycle GAN for head-and-neck radiation therapy. *Comput. Med. Imaging Graph.* **2021**, *91*, 101953. [[CrossRef](#)] [[PubMed](#)]
136. Ostankovich, V.; Yagfarov, R.; Rassabin, M.; Gafurov, S. Application of cycleGAN-based augmentation for autonomous driving at night. In Proceedings of the International Conference Nonlinearity, Information and Robotics (NIR), Innopolis, Russia, 3–6 December 2020; pp. 1–5.
137. Huisman, M.; Van Rijn, J.N.; Plaat, A. A survey of deep meta-learning. *Artif. Intell. Rev.* **2021**, *54*, 4483–4541. [[CrossRef](#)]
138. Tian, Y.; Zhao, X.; Huang, W. Meta-learning approaches for learning-to-learn in deep learning: A survey. *Neurocomputing* **2022**, *494*, 203–223. [[CrossRef](#)]
139. Luo, S.; Li, Y.; Gao, P.; Wang, Y.; Serikawa, S. Meta-seg: A survey of meta-learning for image segmentation. *Pattern Recognit.* **2022**, *126*, 108586. [[CrossRef](#)]
140. He, K.; Pu, N.; Lao, M.; Lew, M.S. Few-shot and meta-learning methods for image understanding: A survey. *Int. J. Multimed. Inf. Retr.* **2023**, *12*, 14. [[CrossRef](#)]
141. Jha, A. In the Era of Prompt Learning with Vision-Language Models. *arXiv* **2024**, arXiv:2411.04892.
142. Zhou, K.; Yang, J.; Loy, C.C.; Liu, Z. Learning to prompt for vision-language models. *Int. J. Comput. Vis.* **2022**, *130*, 2337–2348. [[CrossRef](#)]
143. Fang, A.; Ilharco, G.; Wortsman, M.; Wan, Y.; Shankar, V.; Dave, A.; Schmidt, L. Data determines distributional robustness in contrastive language image pre-training (clip). In Proceedings of the International Conference on Machine Learning, Baltimore, MD, USA, 17–23 July 2022; pp. 6216–6234.
144. Li, Y.; Wang, H.; Duan, Y.; Xu, H.; Li, X. Exploring visual interpretability for contrastive language-image pre-training. *arXiv* **2022**, arXiv:2209.07046.
145. Liu, J.; Wang, H.; Yin, W.; Sonke, J.J.; Gavves, E. Click prompt learning with optimal transport for interactive segmentation. In Proceedings of the European Conference on Computer Vision, Milan, Italy, 29 September–4 October 2024; pp. 93–110.
146. Rao, A.; Fisher, A.; Chang, K.; Panagides, J.C.; McNamara, K.; Lee, J.Y.; Aalami, O. IMIL: Interactive Medical Image Learning Framework. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16–22 June 2024; pp. 5241–5250.
147. Li, H.; Liu, H.; Hu, D.; Wang, J.; Oguz, I. Prism: A promptable and robust interactive segmentation model with visual prompts. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Marrakesh, Morocco, 6–10 October 2024; pp. 389–399.
148. Marinov, Z.; Jäger, P.F.; Egger, J.; Kleesiek, J.; Stiefelhagen, R. Deep interactive segmentation of medical images: A systematic review and taxonomy. *IEEE Trans. Pattern Anal. Mach. Intell.* **2024**, *46*, 10998–11018. [[CrossRef](#)] [[PubMed](#)]
149. Jain, P.; Ienco, D.; Interdonato, R.; Berchoux, T.; Marcos, D. SenCLIP: Enhancing zero-shot land-use mapping for Sentinel-2 with ground-level prompting. *arXiv* **2024**, arXiv:2412.08536.
150. Zhao, M.; Li, M.; Peng, S.L.; Li, J. A novel deep learning model compression algorithm. *Electronics* **2022**, *11*, 1066. [[CrossRef](#)]
151. Mohammed, S.B.; Krothapalli, B.; Althath, C. Advanced Techniques for Storage Optimization in Resource-Constrained Systems Using AI and Machine Learning. *J. Sci. Technol.* **2023**, *4*, 89–125.
152. Vadera, S.; Ameen, S. Methods for pruning deep neural networks. *IEEE Access* **2022**, *10*, 63280–63300. [[CrossRef](#)]
153. Cheng, H.; Zhang, M.; Shi, J.Q. A Survey on Deep Neural Network Pruning: Taxonomy, Comparison, Analysis, and Recommendations. *IEEE Trans. Pattern Anal. Mach. Intell.* **2024**, *46*, 10558–10578. [[CrossRef](#)] [[PubMed](#)]

154. Daghero, F.; Pagliari, D.J.; Poncino, M. Energy-efficient deep learning inference on edge devices. In *Advances in Computers*. Academic Press Inc., United States, **2021**, *122*, 247–301.
155. Abdolrasol, M.G.; Hussain, S.S.; Ustun, T.S.; Sarker, M.R.; Hannan, M.A.; Mohamed, R.; Ali, J.A.; Mekhilef, S.; Milad, A. Artificial neural networks based optimization techniques: A review. *Electronics* **2021**, *10*, 2689. [\[CrossRef\]](#)
156. Zhang, W.; Ji, M.; Yu, H.; Zhen, C. ReLP: Reinforcement learning pruning method based on prior knowledge. *Neural Process. Lett.* **2023**, *55*, 4661–4678. [\[CrossRef\]](#)
157. Zakariyya, I.; Kalutarage, H.; Al-Kadri, M.O. Towards a robust, effective and resource efficient machine learning technique for IoT security monitoring. *Comput. Secur.* **2023**, *133*, 103388. [\[CrossRef\]](#)
158. Rokh, B.; Azarpeyvand, A.; Khanteymoori, A. A comprehensive survey on model quantization for deep neural networks in image classification. *ACM Trans. Intell. Syst. Technol.* **2023**, *14*, 1–50. [\[CrossRef\]](#)
159. Qin, H.; Zhang, Y.; Ding, Y.; Liu, X.; Danelljan, M.; Yu, F. QuantSR: Accurate low-bit quantization for efficient image super-resolution. *Adv. Neural Inf. Process. Syst.* **2024**, *36*.
160. Alkhulaifi, A.; Alsahli, F.; Ahmad, I. Knowledge distillation in deep learning and its applications. *PeerJ Comput. Sci.* **2021**, *7*, e474. [\[CrossRef\]](#)
161. Xu, Q.; Li, Y.; Shen, J.; Liu, J.K.; Tang, H.; Pan, G. Constructing deep spiking neural networks from artificial neural networks with knowledge distillation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 7886–7895.
162. Wang, J.; Wu, Y.; Liu, M.; Yang, M.; Liang, H. A real-time trajectory optimization method for hypersonic vehicles based on a deep neural network. *Aerospace* **2022**, *9*, 188. [\[CrossRef\]](#)
163. Zhang, L.; Bao, C.; Ma, K. Self-distillation: Towards efficient and compact neural networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 4388–4403. [\[CrossRef\]](#)
164. Tian, G.; Chen, J.; Zeng, X.; Liu, Y. Pruning by training: A novel deep neural network compression framework for image processing. *IEEE Signal Process. Lett.* **2021**, *28*, 344–348. [\[CrossRef\]](#)
165. Weng, O. Neural network quantization for efficient inference: A survey. *arXiv* **2021**, arXiv:2112.06126.
166. Tang, J.; Shivanna, R.; Zhao, Z.; Lin, D.; Singh, A.; Chi, E.H.; Jain, S. Understanding and improving knowledge distillation. *arXiv* **2020**, arXiv:2002.03532.
167. Luo, S.; Fang, G.; Song, M. Deep semantic image compression via cooperative network pruning. *J. Vis. Commun. Image Represent.* **2023**, *95*, 103897. [\[CrossRef\]](#)
168. Shafiq, M.; Gu, Z. Deep residual learning for image recognition: A survey. *Appl. Sci.* **2022**, *12*, 8972. [\[CrossRef\]](#)
169. Xie, G.; Ren, J.; Marshall, S.; Zhao, H.; Li, R.; Chen, R. Self-attention enhanced deep residual network for spatial image steganalysis. *Digit. Signal Process.* **2023**, *139*, 104063. [\[CrossRef\]](#)
170. Liu, F.; Ren, X.; Zhang, Z.; Sun, X.; Zou, Y. Rethinking skip connection with layer normalization in transformers and resnets. *arXiv* **2021**, arXiv:2105.07205.
171. Shehab, L.H.; Fahmy, O.M.; Gasser, S.M.; El-Mahallawy, M.S. An efficient brain tumor image segmentation based on deep residual networks (ResNets). *J. King Saud Univ.-Eng. Sci.* **2021**, *33*, 404–412. [\[CrossRef\]](#)
172. Alotaibi, B.; Alotaibi, M. A hybrid deep ResNet and inception model for hyperspectral image classification. *PFG-J. Photogramm. Remote Sens. Geoinf. Sci.* **2020**, *88*, 463–476. [\[CrossRef\]](#)
173. Zhang, C.; Benz, P.; Argaw, D.M.; Lee, S.; Kim, J.; Rameau, F.; Bazin, J.C.; Kweon, I.S. Resnet or densenet? introducing dense shortcuts to resnet. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 3–8 January 2021; pp. 3550–3559.
174. Yadav, D.; Jalal, A.; Garlapati, D.; Hossain, K.; Goyal, A.; Pant, G. Deep learning-based ResNeXt model in phycological studies for future. *Algal Res.* **2020**, *50*, 102018. [\[CrossRef\]](#)
175. Hasan, N.; Bao, Y.; Shawon, A.; Huang, Y. DenseNet convolutional neural networks application for predicting COVID-19 using CT image. *SN Comput. Sci.* **2021**, *2*, 389. [\[CrossRef\]](#) [\[PubMed\]](#)
176. LIU, J.w.; LIU, J.w.; LUO, X.l. Research progress in attention mechanism in deep learning. *Chin. J. Eng.* **2021**, *43*, 1499–1511.
177. Ghaffarian, S.; Valente, J.; Van Der Voort, M.; Tekinerdogan, B. Effect of attention mechanism in deep learning-based remote sensing image processing: A systematic literature review. *Remote Sens.* **2021**, *13*, 2965. [\[CrossRef\]](#)
178. Osman, A.A.; Shalaby, M.A.W.; Soliman, M.M.; Elsayed, K.M. A survey on attention-based models for image captioning. *Int. J. Adv. Comput. Sci. Appl.* **2023**, *14*. [\[CrossRef\]](#)
179. Zhao, J.; Hou, X.; Pan, M.; Zhang, H. Attention-based generative adversarial network in medical imaging: A narrative review. *Comput. Biol. Med.* **2022**, *149*, 105948. [\[CrossRef\]](#)
180. Liu, Y.; Shao, Z.; Hoffmann, N. Global attention mechanism: Retain information to enhance channel-spatial interactions. *arXiv* **2021**, arXiv:2112.05561.
181. Li, J.; Yan, Y.; Liao, S.; Yang, X.; Shao, L. Local-to-global self-attention in vision transformers. *arXiv* **2021**, arXiv:2107.04735.

182. Mehrani, P.; Tsotsos, J.K. Self-attention in vision transformers performs perceptual grouping, not attention. *Front. Comput. Sci.* **2023**, *5*, 1178450. [\[CrossRef\]](#)
183. Chen, X.; Pan, J.; Lu, J.; Fan, Z.; Li, H. Hybrid cnn-transformer feature fusion for single image deraining. In Proceedings of the AAAI Conference on Artificial Intelligence, Washington, DC, USA, 7–14 February 2023; Volume 37, pp. 378–386.
184. Sardar, A.S.; Ranjan, V. Enhancing Computer Vision Performance: A Hybrid Deep Learning Approach with CNNs and Vision Transformers. In Proceedings of the International Conference on Computer Vision and Image Processing, Jammu, India, 3–5 November 2023; pp. 591–602.
185. Zhang, Z.; Jiang, Y.; Jiang, J.; Wang, X.; Luo, P.; Gu, J. Star: A structure-aware lightweight transformer for real-time image enhancement. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 11–17 October 2021; pp. 4106–4115.
186. Wang, L.; Chen, W.; Yang, W.; Bi, F.; Yu, F.R. A state-of-the-art review on image synthesis with generative adversarial networks. *IEEE Access* **2020**, *8*, 63514–63537. [\[CrossRef\]](#)
187. Shamsolmoali, P.; Zareapoor, M.; Granger, E.; Zhou, H.; Wang, R.; Celebi, M.E.; Yang, J. Image synthesis with adversarial networks: A comprehensive survey and case studies. *Inf. Fusion* **2021**, *72*, 126–146. [\[CrossRef\]](#)
188. Lee, I.H.; Chung, W.Y.; Park, C.G. Style transformation super-resolution GAN for extremely small infrared target image. *Pattern Recognit. Lett.* **2023**, *174*, 1–9. [\[CrossRef\]](#)
189. Agnese, J.; Herrera, J.; Tao, H.; Zhu, X. A survey and taxonomy of adversarial neural networks for text-to-image synthesis. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2020**, *10*, e1345. [\[CrossRef\]](#)
190. Sharma, P.; Kumar, M.; Sharma, H.K.; Biju, S.M. Generative adversarial networks (GANs): Introduction, Taxonomy, Variants, Limitations, and Applications. *Multimed. Tools Appl.* **2024**, *83*, 88811–88858. [\[CrossRef\]](#)
191. Stanczuk, J.; Etmann, C.; Kreusser, L.M.; Schönlieb, C.B. Wasserstein GANs work because they fail (to approximate the Wasserstein distance). *arXiv* **2021**, arXiv:2103.01678.
192. Raman, G.; Cao, X.; Li, A.; Raman, G.; Peng, J.C.H.; Lu, J. CGANs-based real-time stability region determination for inverter-based systems. In Proceedings of the IEEE Power & Energy Society General Meeting (PESGM), Montreal, QC, Canada, 2–6 August 2020; pp. 1–5.
193. Khanuja, S.S.; Khanuja, H.K. GAN challenges and optimal solutions. *Int. Res. J. Eng. Technol. (IRJET)* **2021**, *8*, 836–840.
194. Biau, G.; Sangnier, M.; Tanielian, U. Some theoretical insights into Wasserstein GANs. *J. Mach. Learn. Res.* **2021**, *22*, 1–45.
195. Ahmad, Z.; Jaffri, Z.u.A.; Chen, M.; Bao, S. Understanding GANs: Fundamentals, variants, training challenges, applications, and open problems. *Multimed. Tools Appl.* **2024**, 1–77. [\[CrossRef\]](#)
196. Li, Z.; Li, D.; Xu, C.; Wang, W.; Hong, Q.; Li, Q.; Tian, J. Tfcns: A cnn-transformer hybrid network for medical image segmentation. In Proceedings of the International Conference on Artificial Neural Networks, Bristol, UK, 6–9 September 2022; pp. 781–792.
197. Zhao, M.; Cao, G.; Huang, X.; Yang, L. Hybrid transformer-CNN for real image denoising. *IEEE Signal Process. Lett.* **2022**, *29*, 1252–1256. [\[CrossRef\]](#)
198. Gupta, D.; Suman, S.; Ekbal, A. Hierarchical deep multi-modal network for medical visual question answering. *Expert Syst. Appl.* **2021**, *164*, 113993. [\[CrossRef\]](#)
199. Liang, Y.; Wang, X.; Duan, X.; Zhu, W. Multi-modal contextual graph neural network for text visual question answering. In Proceedings of the 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2021; pp. 3491–3498.
200. Wang, Y.; Qiu, Y.; Cheng, P.; Zhang, J. Hybrid CNN-transformer features for visual place recognition. *IEEE Trans. Circuits Syst. Video Technol.* **2022**, *33*, 1109–1122. [\[CrossRef\]](#)
201. Weng, W.; Zhang, Y.; Xiong, Z. Event-based video reconstruction using transformer. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 2563–2572.
202. Tang, Q.; Liang, J.; Zhu, F. A comparative review on multi-modal sensors fusion based on deep learning. *Signal Process.* **2023**, *213*, 109165. [\[CrossRef\]](#)
203. Park, S.; Vien, A.G.; Lee, C. Cross-modal transformers for infrared and visible image fusion. *IEEE Trans. Circuits Syst. Video Technol.* **2023**, *34*, 770–785. [\[CrossRef\]](#)
204. He, X.; Wang, Y.; Zhao, S.; Chen, X. Co-attention fusion network for multimodal skin cancer diagnosis. *Pattern Recognit.* **2023**, *133*, 108990. [\[CrossRef\]](#)
205. Xu, L.; Tang, Q.; Zheng, B.; Lv, J.; Li, W.; Zeng, X. CGFTrans: Cross-Modal Global Feature Fusion Transformer for Medical Report Generation. *IEEE J. Biomed. Health Inform.* **2024**, *28*, 5600–5612. [\[CrossRef\]](#) [\[PubMed\]](#)
206. Li, Y.; Ibanez-Guzman, J. Lidar for autonomous driving: The principles, challenges, and trends for automotive lidar and perception systems. *IEEE Signal Process. Mag.* **2020**, *37*, 50–61. [\[CrossRef\]](#)
207. Reinke, A.; Tizabi, M.D.; Sudre, C.H.; Eisenmann, M.; Radsch, T.; Baumgartner, M.; Acion, L.; Antonelli, M.; Arbel, T.; Bakas, S.; et al. Common limitations of image processing metrics: A picture story. *arXiv* **2021**, arXiv:2104.05642.
208. Singh, S.; Mittal, N.; Singh, H. Classification of various image fusion algorithms and their performance evaluation metrics. *Comput. Intell. Mach. Learn. Healthc. Inform.* **2020**, 179–198. [\[CrossRef\]](#)

209. Wang, Z.; Wang, E.; Zhu, Y. Image segmentation evaluation: A survey of methods. *Artif. Intell. Rev.* **2020**, *53*, 5637–5674. [\[CrossRef\]](#)
210. Zhou, J.; Gandomi, A.H.; Chen, F.; Holzinger, A. Evaluating the quality of machine learning explanations: A survey on methods and metrics. *Electronics* **2021**, *10*, 593. [\[CrossRef\]](#)
211. Baraheem, S.S.; Le, T.N.; Nguyen, T.V. Image synthesis: A review of methods, datasets, evaluation metrics, and future outlook. *Artif. Intell. Rev.* **2023**, *56*, 10813–10865. [\[CrossRef\]](#)
212. Luo, G.; Cheng, L.; Jing, C.; Zhao, C.; Song, G. A thorough review of models, evaluation metrics, and datasets on image captioning. *IET Image Process.* **2022**, *16*, 311–332. [\[CrossRef\]](#)
213. Zhou, S.K.; Greenspan, H.; Davatzikos, C.; Duncan, J.S.; Van Ginneken, B.; Madabhushi, A.; Prince, J.L.; Rueckert, D.; Summers, R.M. A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises. *Proc. IEEE* **2021**, *109*, 820–838. [\[CrossRef\]](#) [\[PubMed\]](#)
214. Suganyadevi, S.; Seethalakshmi, V.; Balasamy, K. A review on deep learning in medical image analysis. *Int. J. Multimed. Inf. Retr.* **2022**, *11*, 19–38. [\[CrossRef\]](#)
215. Narin, A.; Kaya, C.; Pamuk, Z. Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks. *Pattern Anal. Appl.* **2021**, *24*, 1207–1220. [\[CrossRef\]](#) [\[PubMed\]](#)
216. Allugunti, V.R. A machine learning model for skin disease classification using convolution neural network. *Int. J. Comput. Program. Database Manag.* **2022**, *3*, 141–147. [\[CrossRef\]](#)
217. Francolini, G.; Desideri, I.; Stocchi, G.; Salvestrini, V.; Ciccone, L.P.; Garlatti, P.; Loi, M.; Livi, L. Artificial Intelligence in radiotherapy: State of the art and future directions. *Med. Oncol.* **2020**, *37*, 1–9. [\[CrossRef\]](#)
218. Bera, K.; Braman, N.; Gupta, A.; Velcheti, V.; Madabhushi, A. Predicting cancer outcomes with radiomics and artificial intelligence in radiology. *Nat. Rev. Clin. Oncol.* **2022**, *19*, 132–146. [\[CrossRef\]](#)
219. Ebrahimi, A.; Luo, S.; Disease Neuroimaging Initiative, f.t.A. Convolutional neural networks for Alzheimer’s disease detection on MRI images. *J. Med. Imaging* **2021**, *8*, 024503. [\[CrossRef\]](#)
220. Hatuwal, B.K.; Thapa, H.C. Lung cancer detection using convolutional neural network on histopathological images. *Int. J. Comput. Trends Technol.* **2020**, *68*, 21–24. [\[CrossRef\]](#)
221. Samanta, A.; Saha, A.; Satapathy, S.C.; Fernandes, S.L.; Zhang, Y.D. Automated detection of diabetic retinopathy using convolutional neural networks on a small dataset. *Pattern Recognit. Lett.* **2020**, *135*, 293–298. [\[CrossRef\]](#)
222. Krishnan, R.; Rajpurkar, P.; Topol, E.J. Self-supervised learning in medicine and healthcare. *Nat. Biomed. Eng.* **2022**, *6*, 1346–1352. [\[CrossRef\]](#) [\[PubMed\]](#)
223. Huang, S.C.; Pareek, A.; Jensen, M.; Lungren, M.P.; Yeung, S.; Chaudhari, A.S. Self-supervised learning for medical image classification: A systematic review and implementation guidelines. *npj Digit. Med.* **2023**, *6*, 74. [\[CrossRef\]](#) [\[PubMed\]](#)
224. Shurrab, S.; Duwairi, R. Self-supervised learning methods and applications in medical imaging analysis: A survey. *PeerJ Comput. Sci.* **2022**, *8*, e1045. [\[CrossRef\]](#) [\[PubMed\]](#)
225. Celi, L.A.; Cellini, J.; Charpignon, M.L.; Dee, E.C.; Dernoncourt, F.; Eber, R.; Mitchell, W.G.; Moukheiber, L.; Schirmer, J.; Situ, J.; et al. Sources of bias in artificial intelligence that perpetuate healthcare disparities—A global review. *PLoS Digit. Health* **2022**, *1*, e0000022. [\[CrossRef\]](#) [\[PubMed\]](#)
226. Chowdhury, R.H. Intelligent systems for healthcare diagnostics and treatment. *World J. Adv. Res. Rev.* **2024**, *23*, 007–015. [\[CrossRef\]](#)
227. Xie, Y.; Lu, L.; Gao, F.; He, S.j.; Zhao, H.j.; Fang, Y.; Yang, J.m.; An, Y.; Ye, Z.w.; Dong, Z. Integration of artificial intelligence, blockchain, and wearable technology for chronic disease management: A new paradigm in smart healthcare. *Curr. Med. Sci.* **2021**, *41*, 1123–1133. [\[CrossRef\]](#) [\[PubMed\]](#)
228. Chawla, N. AI, IOT and Wearable Technology for Smart Healthcare—A Review. *Int. J. Recent Res. Asp.* **2020**, *7*, 9–14.
229. Kuutti, S.; Bowden, R.; Jin, Y.; Barber, P.; Fallah, S. A survey of deep learning applications to autonomous vehicle control. *IEEE Trans. Intell. Transp. Syst.* **2020**, *22*, 712–733. [\[CrossRef\]](#)
230. Grigorescu, S.; Trasnea, B.; Cocias, T.; Macesanu, G. A survey of deep learning techniques for autonomous driving. *J. Field Robot.* **2020**, *37*, 362–386. [\[CrossRef\]](#)
231. Tran, L.A.; Do, T.D.; Park, D.C.; Le, M.H. Enhancement of robustness in object detection module for advanced driver assistance systems. In Proceedings of the International Conference on System Science and Engineering (ICSSE), Nha Trang, Vietnam, 26–28 August 2021; pp. 158–163.
232. Farooq, M.A.; Corcoran, P.; Rotariu, C.; Shariff, W. Object detection in thermal spectrum for advanced driver-assistance systems (ADAS). *IEEE Access* **2021**, *9*, 156465–156481. [\[CrossRef\]](#)
233. Tran, L.A.; Do, T.D.; Park, D.C.; Le, M.H. Robustness Enhancement of Object Detection in Advanced Driver Assistance Systems (ADAS). *arXiv* **2021**, arXiv:2105.01580.
234. Li, G.; Li, S.; Li, S.; Qin, Y.; Cao, D.; Qu, X.; Cheng, B. Deep reinforcement learning enabled decision-making for autonomous driving at intersections. *Automot. Innov.* **2020**, *3*, 374–385. [\[CrossRef\]](#)

235. Harrison, K.; Ingole, R.; Surabhi, S.N.R.D. Enhancing Autonomous Driving: Evaluations Of AI And ML Algorithms. *Educ. Adm. Theory Pract.* **2024**, *30*, 4117–4126. [\[CrossRef\]](#)
236. Jeyaraman, J.; Malaiyappan, J.N.A.; Sistla, S.M.K. Advancements in Reinforcement Learning Algorithms for Autonomous Systems. *Int. J. Innov. Sci. Res. Technol. (IJISRT)* **2024**, *9*, 1941–1946.
237. Ekatpure, R. Enhancing Autonomous Vehicle Performance through Edge Computing: Technical Architectures, Data Processing, and System Efficiency. *Appl. Res. Artif. Intell. Cloud Comput.* **2023**, *6*, 17–34.
238. Lv, Z.; Chen, D.; Wang, Q. Diversified technologies in internet of vehicles under intelligent edge computing. *IEEE Trans. Intell. Transp. Syst.* **2020**, *22*, 2048–2059. [\[CrossRef\]](#)
239. Ma, Y.; Wang, Z.; Yang, H.; Yang, L. Artificial intelligence applications in the development of autonomous vehicles: A survey. *IEEE/CAA J. Autom. Sin.* **2020**, *7*, 315–329. [\[CrossRef\]](#)
240. Bathla, G.; Bhadane, K.; Singh, R.K.; Kumar, R.; Aluvalu, R.; Krishnamurthi, R.; Kumar, A.; Thakur, R.; Basheer, S. Autonomous vehicles and intelligent automation: Applications, challenges, and opportunities. *Mob. Inf. Syst.* **2022**, *2022*, 7632892. [\[CrossRef\]](#)
241. Yuan, Q.; Shen, H.; Li, T.; Li, Z.; Li, S.; Jiang, Y.; Xu, H.; Tan, W.; Yang, Q.; Wang, J.; et al. Deep learning in environmental remote sensing: Achievements and challenges. *Remote Sens. Environ.* **2020**, *241*, 111716. [\[CrossRef\]](#)
242. Li, J.; Pei, Y.; Zhao, S.; Xiao, R.; Sang, X.; Zhang, C. A review of remote sensing for environmental monitoring in China. *Remote Sens.* **2020**, *12*, 1130. [\[CrossRef\]](#)
243. Chen, J.; Chen, S.; Fu, R.; Li, D.; Jiang, H.; Wang, C.; Peng, Y.; Jia, K.; Hicks, B.J. Remote sensing big data for water environment monitoring: Current status, challenges, and future prospects. *Earth's Future* **2022**, *10*, e2021EF002289. [\[CrossRef\]](#)
244. Pi, Y.; Nath, N.D.; Behzadan, A.H. Convolutional neural networks for object detection in aerial imagery for disaster response and recovery. *Adv. Eng. Inform.* **2020**, *43*, 101009. [\[CrossRef\]](#)
245. Park, J.; Lee, D.; Lee, J.; Cheon, E.; Jeong, H. Study on Disaster Response Strategies Using Multi-Sensors Satellite Imagery. *Korean J. Remote Sens.* **2023**, *39*, 755–770.
246. Rashid, M.; Bari, B.S.; Yusup, Y.; Kamaruddin, M.A.; Khan, N. A comprehensive review of crop yield prediction using machine learning approaches with special emphasis on palm oil yield prediction. *IEEE Access* **2021**, *9*, 63406–63439. [\[CrossRef\]](#)
247. Masolele, R.N.; De Sy, V.; Herold, M.; Marcos, D.; Verbesselt, J.; Gieseke, F.; Mullissa, A.G.; Martius, C. Spatial and temporal deep learning methods for deriving land-use following deforestation: A pan-tropical case study using Landsat time series. *Remote Sens. Environ.* **2021**, *264*, 112600. [\[CrossRef\]](#)
248. Sun, X.; Zhang, Y.; Shi, K.; Zhang, Y.; Li, N.; Wang, W.; Huang, X.; Qin, B. Monitoring water quality using proximal remote sensing technology. *Sci. Total Environ.* **2022**, *803*, 149805. [\[CrossRef\]](#) [\[PubMed\]](#)
249. Shafique, A.; Cao, G.; Khan, Z.; Asad, M.; Aslam, M. Deep learning-based change detection in remote sensing images: A review. *Remote Sens.* **2022**, *14*, 871. [\[CrossRef\]](#)
250. Desai, S.; Ghose, D. Active learning for improved semi-supervised semantic segmentation in satellite images. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 3–8 January 2022; pp. 553–563.
251. Gu, X.; Angelov, P.P.; Zhang, C.; Atkinson, P.M. A semi-supervised deep rule-based approach for complex satellite sensor image analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *44*, 2281–2292. [\[CrossRef\]](#)
252. Raghavan, R.; Verma, D.C.; Pandey, D.; Anand, R.; Pandey, B.K.; Singh, H. Optimized building extraction from high-resolution satellite imagery using deep learning. *Multimed. Tools Appl.* **2022**, *81*, 42309–42323. [\[CrossRef\]](#)
253. Qin, R.; Liu, T. A review of landcover classification with very-high resolution remotely sensed optical images—Analysis unit, model scalability and transferability. *Remote Sens.* **2022**, *14*, 646. [\[CrossRef\]](#)
254. Rezaee, K.; Rezakhani, S.M.; Khosravi, M.R.; Moghimi, M.K. A survey on deep learning-based real-time crowd anomaly detection for secure distributed video surveillance. *Pers. Ubiquitous Comput.* **2024**, *28*, 135–151. [\[CrossRef\]](#)
255. Iqbal, M.J.; Iqbal, M.M.; Ahmad, I.; Alassafi, M.O.; Alfakeeh, A.S.; Alhomoud, A. Real-Time Surveillance Using Deep Learning. *Secur. Commun. Netw.* **2021**, *2021*, 6184756. [\[CrossRef\]](#)
256. Schuartz, F.C.; Fonseca, M.; Munaretto, A. Improving threat detection in networks using deep learning. *Ann. Telecommun.* **2020**, *75*, 133–142. [\[CrossRef\]](#)
257. Raut, M.; Dhavale, S.; Singh, A.; Mehra, A. Insider threat detection using deep learning: A review. In Proceedings of the 3rd International Conference on Intelligent Sustainable Systems (ICISS), Thoothukudi, India, 3–5 Decembe 2020; pp. 856–863.
258. Maddireddy, B.R.; Maddireddy, B.R. Advancing Threat Detection: Utilizing Deep Learning Models for Enhanced Cybersecurity Protocols. *Rev. Esp. Doc. Cient.* **2024**, *18*, 325–355.
259. Salama AbdELminaam, D.; Almansori, A.M.; Taha, M.; Badr, E. A deep facial recognition system using computational intelligent algorithms. *PLoS ONE* **2020**, *15*, e0242269. [\[CrossRef\]](#) [\[PubMed\]](#)
260. Singh, A.; Bhatt, S.; Nayak, V.; Shah, M. Automation of surveillance systems using deep learning and facial recognition. *Int. J. Syst. Assur. Eng. Manag.* **2023**, *14*, 236–245. [\[CrossRef\]](#)
261. Saheb, T. Ethically contentious aspects of artificial intelligence surveillance: A social science perspective. *AI Ethics* **2023**, *3*, 369–379. [\[CrossRef\]](#) [\[PubMed\]](#)

262. Wang, X.; Wu, Y.C.; Zhou, M.; Fu, H. Beyond surveillance: Privacy, ethics, and regulations in face recognition technology. *Front. Big Data* **2024**, *7*, 1337465. [\[CrossRef\]](#)
263. Smith, M.; Miller, S. The ethical application of biometric facial recognition technology. *AI Soc.* **2022**, *37*, 167–175. [\[CrossRef\]](#) [\[PubMed\]](#)
264. Andrejevic, M.; Selwyn, N. Facial recognition technology in schools: Critical questions and concerns. *Learn. Media Technol.* **2020**, *45*, 115–128. [\[CrossRef\]](#)
265. Ferrer, X.; Van Nuenen, T.; Such, J.M.; Coté, M.; Criado, N. Bias and discrimination in AI: A cross-disciplinary perspective. *IEEE Technol. Soc. Mag.* **2021**, *40*, 72–80. [\[CrossRef\]](#)
266. Ntoutsis, E.; Fafalios, P.; Gadiraju, U.; Iosifidis, V.; Nejdil, W.; Vidal, M.E.; Ruggieri, S.; Turini, F.; Papadopoulos, S.; Krasanakis, E.; et al. Bias in data-driven artificial intelligence systems—An introductory survey. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2020**, *10*, e1356. [\[CrossRef\]](#)
267. Lee, R.S.; Lee, R.S. AI ethics, security and privacy. In *Artificial Intelligence in Daily Life*; Springer: Singapore, 2020; pp. 369–384. [\[CrossRef\]](#)
268. Gupta, V.; Sambyal, N.; Sharma, A.; Kumar, P. Restoration of artwork using deep neural networks. *Evol. Syst.* **2021**, *12*, 439–446. [\[CrossRef\]](#)
269. Gaber, J.A.; Youssef, S.M.; Fathalla, K.M. The role of artificial intelligence and machine learning in preserving cultural heritage and art works via virtual restoration. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2023**, *10*, 185–190. [\[CrossRef\]](#)
270. Mendoza, M.A.D.; De La Hoz Franco, E.; Gómez, J.E.G. Technologies for the preservation of cultural heritage—A systematic review of the literature. *Sustainability* **2023**, *15*, 1059. [\[CrossRef\]](#)
271. Trček, D. Cultural heritage preservation by using blockchain technologies. *Herit. Sci.* **2022**, *10*, 6. [\[CrossRef\]](#)
272. Belhi, A.; Bouras, A.; Al-Ali, A.K.; Fofou, S. A machine learning framework for enhancing digital experiences in cultural heritage. *J. Enterp. Inf. Manag.* **2023**, *36*, 734–746. [\[CrossRef\]](#)
273. Leshkevich, T.; Motozhanets, A. Social perception of artificial intelligence and digitization of cultural heritage: Russian context. *Appl. Sci.* **2022**, *12*, 2712. [\[CrossRef\]](#)
274. Yu, T.; Lin, C.; Zhang, S.; Wang, C.; Ding, X.; An, H.; Liu, X.; Qu, T.; Wan, L.; You, S.; et al. Artificial intelligence for Dunhuang cultural heritage protection: the project and the dataset. *Int. J. Comput. Vis.* **2022**, *130*, 2646–2673. [\[CrossRef\]](#)
275. Fiorucci, M.; Khoroshiltseva, M.; Pontil, M.; Traviglia, A.; Del Bue, A.; James, S. Machine learning for cultural heritage: A survey. *Pattern Recognit. Lett.* **2020**, *133*, 102–108. [\[CrossRef\]](#)
276. Kusters, R.; Misevic, D.; Berry, H.; Cully, A.; Le Cunff, Y.; Dandoy, L.; Díaz-Rodríguez, N.; Ficher, M.; Grizou, J.; Othmani, A.; et al. Interdisciplinary research in artificial intelligence: Challenges and opportunities. *Front. Big Data* **2020**, *3*, 577974. [\[CrossRef\]](#)
277. Meron, Y. Graphic Design and Artificial Intelligence: Interdisciplinary Challenges for Designers in the Search for Research Collaboration. In Proceedings of the DRS Conference Proceedings, Bilbao, Spain, 25 June–3 July 2022. [\[CrossRef\]](#)
278. Audry, S. *Art in the Age of Machine Learning*; MIT Press: Cambridge, MA, USA, 2021.
279. Mello, M.M.; Wang, C.J. Ethics and governance for digital disease surveillance. *Science* **2020**, *368*, 951–954. [\[CrossRef\]](#) [\[PubMed\]](#)
280. Dhirani, L.L.; Mukhtiar, N.; Chowdhry, B.S.; Newe, T. Ethical dilemmas and privacy issues in emerging technologies: A review. *Sensors* **2023**, *23*, 1151. [\[CrossRef\]](#)
281. Drukker, K.; Chen, W.; Gichoya, J.; Grusauskas, N.; Kalpathy-Cramer, J.; Koyejo, S.; Myers, K.; Sá, R.C.; Sahiner, B.; Whitney, H.; et al. Toward fairness in artificial intelligence for medical image analysis: Identification and mitigation of potential biases in the roadmap from data collection to model deployment. *J. Med. Imaging* **2023**, *10*, 061104. [\[CrossRef\]](#) [\[PubMed\]](#)
282. Tripathi, S.; Musiolik, T.H. Fairness and ethics in artificial intelligence-based medical imaging. In *Ethical Implications of Reshaping Healthcare with Emerging Technologies*; IGI Global: Hershey, PA, USA, 2022; pp. 71–85.
283. Santosh, K.; Gaur, L. *Artificial Intelligence and Machine Learning in Public Healthcare: Opportunities and Societal Impact*; Springer: Singapore, 2022.
284. Panigutti, C.; Monreale, A.; Comandè, G.; Pedreschi, D. Ethical, societal and legal issues in deep learning for healthcare. *Deep. Learn. Biol. Med.* **2022**, 265–313. [\[CrossRef\]](#)
285. Hussain, I.; Nazir, M.B. Empowering Healthcare: AI, ML, and Deep Learning Innovations for Brain and Heart Health. *Int. J. Adv. Eng. Technol. Innov.* **2024**, *1*, 167–188.
286. Khanna, S.; Srivastava, S. Patient-centric ethical frameworks for privacy, transparency, and bias awareness in deep learning-based medical systems. *Appl. Res. Artif. Intell. Cloud Comput.* **2020**, *3*, 16–35.
287. Hogenhout, L. A framework for ethical AI at the United Nations. *arXiv* **2021**, arXiv:2104.12547.
288. Vegesna, V.V. Privacy-Preserving Techniques in AI-Powered Cyber Security: Challenges and Opportunities. *Int. J. Mach. Learn. Sustain. Dev.* **2023**, *5*, 1–8.
289. Dhinakaran, D.; Sankar, S.; Selvaraj, D.; Raja, S.E. Privacy-Preserving Data in IoT-based Cloud Systems: A Comprehensive Survey with AI Integration. *arXiv* **2024**, arXiv:2401.00794.

290. Shanmugam, L.; Tillu, R.; Jangoan, S. Privacy-Preserving AI/ML Application Architectures: Techniques, Trade-offs, and Case Studies. *J. Knowl. Learn. Sci. Technol.* **2023**, *2*, 398–420. [\[CrossRef\]](#)
291. Memarian, B.; Doleck, T. Fairness, Accountability, Transparency, and Ethics (FATE) in Artificial Intelligence (AI), and higher education: A systematic review. *Comput. Educ. Artif. Intell.* **2023**, *5*, 100152. [\[CrossRef\]](#)
292. Akinrinola, O.; Okoye, C.C.; Ofodile, O.C.; Ugochukwu, C.E. Navigating and reviewing ethical dilemmas in AI development: Strategies for transparency, fairness, and accountability. *GSC Adv. Res. Rev.* **2024**, *18*, 050–058. [\[CrossRef\]](#)
293. Lepore, D.; Dolui, K.; Tomashchuk, O.; Shim, H.; Puri, C.; Li, Y.; Chen, N.; Spigarelli, F. Interdisciplinary research unlocking innovative solutions in healthcare. *Technovation* **2023**, *120*, 102511. [\[CrossRef\]](#)
294. Rasheed, K.; Qayyum, A.; Ghaly, M.; Al-Fuqaha, A.; Razi, A.; Qadir, J. Explainable, trustworthy, and ethical machine learning for healthcare: A survey. *Comput. Biol. Med.* **2022**, *149*, 106043. [\[CrossRef\]](#)
295. Geroski, T.; Filipović, N. Artificial Intelligence Empowering Medical Image Processing. In *In Silico Clinical Trials for Cardiovascular Disease: A Finite Element and Machine Learning Approach*; Springer: Cham, Switzerland, 2024; pp. 179–208.
296. Castiglioni, I.; Rundo, L.; Codari, M.; Di Leo, G.; Salvatore, C.; Interlenghi, M.; Gallivanone, F.; Cozzi, A.; D’Amico, N.C.; Sardanelli, F. AI applications to medical images: From machine learning to deep learning. *Phys. Medica* **2021**, *83*, 9–24. [\[CrossRef\]](#)
297. Gupta, S.; Kumar, S.; Chang, K.; Lu, C.; Singh, P.; Kalpathy-Cramer, J. Collaborative privacy-preserving approaches for distributed deep learning using multi-institutional data. *RadioGraphics* **2023**, *43*, e220107. [\[CrossRef\]](#)
298. Kim, J.C.; Chung, K. Hybrid multi-modal deep learning using collaborative concat layer in health bigdata. *IEEE Access* **2020**, *8*, 192469–192480. [\[CrossRef\]](#)
299. Qian, Y. Network Science, Big Data Analytics, and Deep Learning: An Interdisciplinary Approach to the Study of Citation, Social and Collaboration Networks. Ph.D. Thesis, Queen Mary University of London, London, UK, 2021.
300. Peters, D.; Vold, K.; Robinson, D.; Calvo, R.A. Responsible AI—Two frameworks for ethical design practice. *IEEE Trans. Technol. Soc.* **2020**, *1*, 34–47. [\[CrossRef\]](#)
301. Rakova, B.; Yang, J.; Cramer, H.; Chowdhury, R. Where responsible AI meets reality: Practitioner perspectives on enablers for shifting organizational practices. *Proc. ACM Hum.-Comput. Interact.* **2021**, *5*, 1–23. [\[CrossRef\]](#)
302. Sarker, I.; Colman, A.; Han, J.; Watters, P. *Context-Aware Machine Learning and Mobile Data Analytics: Automated Rule-Based Services with Intelligent Decision-Making*; Springer: Cham, Switzerland, 2021.
303. Unger, M.; Tuzhilin, A.; Livne, A. Context-aware recommendations based on deep learning frameworks. *ACM Trans. Manag. Inf. Syst. (TMIS)* **2020**, *11*, 1–15. [\[CrossRef\]](#)
304. Jeong, S.Y.; Kim, Y.K. Deep learning-based context-aware recommender system considering contextual features. *Appl. Sci.* **2021**, *12*, 45. [\[CrossRef\]](#)
305. Bansal, M.A.; Sharma, D.R.; Kathuria, D.M. A systematic review on data scarcity problem in deep learning: solution and applications. *ACM Comput. Surv. (CSUR)* **2022**, *54*, 1–29. [\[CrossRef\]](#)
306. Alzubaidi, L.; Bai, J.; Al-Sabaawi, A.; Santamaría, J.; Albahri, A.S.; Al-dabbagh, B.S.N.; Fadhel, M.A.; Manoufali, M.; Zhang, J.; Al-Timemy, A.H.; et al. A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications. *J. Big Data* **2023**, *10*, 46. [\[CrossRef\]](#)
307. Dewi, C.; Chen, R.C.; Liu, Y.T.; Tai, S.K. Synthetic Data generation using DCGAN for improved traffic sign recognition. *Neural Comput. Appl.* **2022**, *34*, 21465–21480. [\[CrossRef\]](#)
308. de Melo, C.M.; Torralba, A.; Guibas, L.; DiCarlo, J.; Chellappa, R.; Hodgins, J. Next-generation deep learning based on simulators and synthetic data. *Trends Cogn. Sci.* **2022**, *26*, 174–187. [\[CrossRef\]](#) [\[PubMed\]](#)
309. Wen, Q.; Sun, L.; Yang, F.; Song, X.; Gao, J.; Wang, X.; Xu, H. Time series data augmentation for deep learning: A survey. *arXiv* **2020**, arXiv:2002.12478.
310. Khosla, C.; Saini, B.S. Enhancing performance of deep learning models with different data augmentation techniques: A survey. In Proceedings of the International Conference on Intelligent Engineering and Management (ICIEM), London, UK, 17–19 June 2020; pp. 79–85.
311. Wani, M.A.; Bhat, F.A.; Afzal, S.; Khan, A.I. *Advances in Deep Learning*; Springer: Singapore, 2020. [\[CrossRef\]](#)
312. Freire, P.; Srivallapanondh, S.; Napoli, A.; Prilepsky, J.E.; Turitsyn, S.K. Computational complexity evaluation of neural network applications in signal processing. *arXiv* **2022**, arXiv:2206.12191.
313. Murshed, M.S.; Murphy, C.; Hou, D.; Khan, N.; Ananthanarayanan, G.; Hussain, F. Machine learning at the network edge: A survey. *ACM Comput. Surv. (CSUR)* **2021**, *54*, 1–37. [\[CrossRef\]](#)
314. Merenda, M.; Porcaro, C.; Iero, D. Edge machine learning for ai-enabled iot devices: A review. *Sensors* **2020**, *20*, 2533. [\[CrossRef\]](#)
315. Acun, B.; Murphy, M.; Wang, X.; Nie, J.; Wu, C.J.; Hazelwood, K. Understanding training efficiency of deep learning recommendation models at scale. In Proceedings of the IEEE International Symposium on High-Performance Computer Architecture (HPCA), Seoul, Republic of Korea, 27 February–3 March 2021; pp. 802–814.
316. Menghani, G. Efficient deep learning: A survey on making deep learning models smaller, faster, and better. *ACM Comput. Surv.* **2023**, *55*, 1–37. [\[CrossRef\]](#)

317. Stiglic, G.; Kocbek, P.; Fijacko, N.; Zitnik, M.; Verbert, K.; Cilar, L. Interpretability of machine learning-based prediction models in healthcare. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2020**, *10*, e1379. [[CrossRef](#)]
318. Brigo, D.; Huang, X.; Pallavicini, A.; Borde, H.S.d.O. Interpretability in deep learning for finance: A case study for the Heston model. *arXiv* **2021**, arXiv:2104.09476. [[CrossRef](#)]
319. Von Eschenbach, W.J. Transparency and the black box problem: Why we do not trust AI. *Philos. Technol.* **2021**, *34*, 1607–1622. [[CrossRef](#)]
320. Franzoni, V. From black box to glass box: Advancing transparency in artificial intelligence systems for ethical and trustworthy AI. In Proceedings of the International Conference on Computational Science and Its Applications, Athens, Greece, 3–6 July 2023; pp. 118–130.
321. Saisubramanian, S.; Galhotra, S.; Zilberstein, S. Balancing the tradeoff between clustering value and interpretability. In Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, New York, NY, USA, 7–9 February 2020; pp. 351–357.
322. He, C.; Ma, M.; Wang, P. Extract interpretability-accuracy balanced rules from artificial neural networks: A review. *Neurocomputing* **2020**, *387*, 346–358. [[CrossRef](#)]
323. Zhao, L.; Liu, T.; Peng, X.; Metaxas, D. Maximum-entropy adversarial data augmentation for improved generalization and robustness. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 14435–14447.
324. Zhang, L.; Deng, Z.; Kawaguchi, K.; Ghorbani, A.; Zou, J. How does mixup help with robustness and generalization? *arXiv* **2020**, arXiv:2010.04819.
325. Bai, T.; Luo, J.; Zhao, J.; Wen, B.; Wang, Q. Recent advances in adversarial training for adversarial robustness. *arXiv* **2021**, arXiv:2102.01356.
326. Han, D.; Wang, Z.; Zhong, Y.; Chen, W.; Yang, J.; Lu, S.; Shi, X.; Yin, X. Evaluating and improving adversarial robustness of machine learning-based network intrusion detectors. *IEEE J. Sel. Areas Commun.* **2021**, *39*, 2632–2647. [[CrossRef](#)]
327. Taori, R.; Dave, A.; Shankar, V.; Carlini, N.; Recht, B.; Schmidt, L. Measuring robustness to natural distribution shifts in image classification. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 18583–18599.
328. Wiles, O.; Goyal, S.; Stimberg, F.; Alvisé-Rebuffi, S.; Ktena, I.; Dvijotham, K.; Cemgil, T. A fine-grained analysis on distribution shift. *arXiv* **2021**, arXiv:2110.11328.
329. Puyol-Antón, E.; Ruijsink, B.; Piechnik, S.K.; Neubauer, S.; Petersen, S.E.; Razavi, R.; King, A.P. Fairness in cardiac MR image analysis: An investigation of bias due to data imbalance in deep learning based segmentation. In Proceedings of the 24th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), Strasbourg, France, 27 September–1 October 2021; pp. 413–423.
330. Shah, M.; Sureja, N. A Comprehensive Review of Bias in Deep Learning Models: Methods, Impacts, and Future Directions. *Arch. Comput. Methods Eng.* **2024**, *32*, 255–267. [[CrossRef](#)]
331. Almeida, D.; Shmarko, K.; Lomas, E. The ethics of facial recognition technologies, surveillance, and accountability in an age of artificial intelligence: A comparative analysis of US, EU, and UK regulatory frameworks. *AI Ethics* **2022**, *2*, 377–387. [[CrossRef](#)]
332. Fontes, C.; Perrone, C. *Ethics of Surveillance: Harnessing the Use of Live Facial Recognition Technologies in Public Spaces for Law Enforcement*; Technical University of Munich: Munich, Germany, 2021.
333. Alikhademi, K.; Drobinina, E.; Prioleau, D.; Richardson, B.; Purves, D.; Gilbert, J.E. A review of predictive policing from the perspective of fairness. *Artif. Intell. Law* **2022**, *30*, 1–17. [[CrossRef](#)]
334. Yen, C.P.; Hung, T.W. Achieving equity with predictive policing algorithms: A social safety net perspective. *Sci. Eng. Ethics* **2021**, *27*, 1–16. [[CrossRef](#)] [[PubMed](#)]
335. Akrim, A.; Gogu, C.; Vingerhoeds, R.; Salaün, M. Self-Supervised Learning for data scarcity in a fatigue damage prognostic problem. *Eng. Appl. Artif. Intell.* **2023**, *120*, 105837. [[CrossRef](#)]
336. Wittscher, L.; Pigorsch, C. Exploring Self-supervised Capsule Networks for Improved Classification with Data Scarcity. In Proceedings of the International Conference on Image Processing and Capsule Networks, Bangkok, Thailand, 20–21 May 2022; pp. 36–50.
337. Bekker, J.; Davis, J. Learning from positive and unlabeled data: A survey. *Mach. Learn.* **2020**, *109*, 719–760. [[CrossRef](#)]
338. Guo, L.Z.; Zhang, Z.Y.; Jiang, Y.; Li, Y.F.; Zhou, Z.H. Safe deep semi-supervised learning for unseen-class unlabeled data. In Proceedings of the International Conference on Machine Learning, Vienna, Austria, 12–18 July 2020; pp. 3897–3906.
339. Huang, W.; Yi, M.; Zhao, X.; Jiang, Z. Towards the generalization of contrastive self-supervised learning. *arXiv* **2021**, arXiv:2111.00743.
340. Kim, D.; Yoo, Y.; Park, S.; Kim, J.; Lee, J. Selfreg: Self-supervised contrastive regularization for domain generalization. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 9619–9628.
341. Wang, D.; Li, M.; Gong, C.; Chandra, V. Attentionas: Improving neural architecture search via attentive sampling. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 6418–6427.

342. White, C.; Zela, A.; Ru, R.; Liu, Y.; Hutter, F. How powerful are performance predictors in neural architecture search? *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 28454–28469.
343. Kim, J.; Chang, S.; Kwak, N. PQK: Model compression via pruning, quantization, and knowledge distillation. *arXiv* **2021**, arXiv:2106.14681.
344. Liang, T.; Glossner, J.; Wang, L.; Shi, S.; Zhang, X. Pruning and quantization for deep neural network acceleration: A survey. *Neurocomputing* **2021**, *461*, 370–403. [[CrossRef](#)]
345. Marković, D.; Grollier, J. Quantum neuromorphic computing. *Appl. Phys. Lett.* **2020**, *117*, 150501. [[CrossRef](#)]
346. Ghosh, S.; Nakajima, K.; Krisnanda, T.; Fujii, K.; Liew, T.C. Quantum neuromorphic computing with reservoir computing networks. *Adv. Quantum Technol.* **2021**, *4*, 2100053. [[CrossRef](#)]
347. Bento, V.; Kohler, M.; Diaz, P.; Mendoza, L.; Pacheco, M.A. Improving deep learning performance by using Explainable Artificial Intelligence (XAI) approaches. *Discov. Artif. Intell.* **2021**, *1*, 1–11. [[CrossRef](#)]
348. Arrieta, A.B.; Díaz-Rodríguez, N.; Del Ser, J.; Bennetot, A.; Tabik, S.; Barbado, A.; García, S.; Gil-López, S.; Molina, D.; Benjamins, R.; et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf. Fusion* **2020**, *58*, 82–115. [[CrossRef](#)]
349. Van der Velden, B.H.; Kuijff, H.J.; Gilhuijs, K.G.; Viergever, M.A. Explainable artificial intelligence (XAI) in deep learning-based medical image analysis. *Med. Image Anal.* **2022**, *79*, 102470. [[CrossRef](#)] [[PubMed](#)]
350. Chen, Z.; Xiao, F.; Guo, F.; Yan, J. Interpretable machine learning for building energy management: A state-of-the-art review. *Adv. Appl. Energy* **2023**, *9*, 100123. [[CrossRef](#)]
351. Molnar, C.; Casalicchio, G.; Bischl, B. Interpretable machine learning—a brief history, state-of-the-art and challenges. In Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Ghent, Belgium, 14–18 September 2020; pp. 417–431.
352. Nannini, L.; Balayn, A.; Smith, A.L. Explainability in AI policies: A critical review of communications, reports, regulations, and standards in the EU, US, and UK. In Proceedings of the ACM Conference on Fairness, Accountability, and Transparency, Chicago, IL, USA, 12–15 June 2023; pp. 1198–1212.
353. Ebers, M. Regulating explainable AI in the European Union. An overview of the current legal framework(s). In *Nordic Yearbook of Law and Informatics*; The Swedish Law and Informatics Research Institute: Stockholm, Swedish, 2020.
354. Alchieri, L.; Badalotti, D.; Bonardi, P.; Bianco, S. An introduction to quantum machine learning: From quantum logic to quantum deep learning. *Quantum Mach. Intell.* **2021**, *3*, 28. [[CrossRef](#)]
355. Peral-García, D.; Cruz-Benito, J. and García-Peñalvo, F.J. Systematic literature review: Quantum machine learning and its applications. *Comput. Sci. Rev.* **2024**, *51*, 100619. [[CrossRef](#)]
356. Dou, W.; Zhao, X.; Yin, X.; Wang, H.; Luo, Y.; Qi, L. Edge computing-enabled deep learning for real-time video optimization in IIoT. *IEEE Trans. Ind. Inform.* **2020**, *17*, 2842–2851. [[CrossRef](#)]
357. Wang, F.; Zhang, M.; Wang, X.; Ma, X.; Liu, J. Deep learning for edge computing applications: A state-of-the-art survey. *IEEE Access* **2020**, *8*, 58322–58336. [[CrossRef](#)]
358. Zhang, C.; Wang, J.; Yen, G.G.; Zhao, C.; Sun, Q.; Tang, Y.; Qian, F.; Kurths, J. When autonomous systems meet accuracy and transferability through AI: A survey. *Patterns* **2020**, *1*, 100050. [[CrossRef](#)]
359. Sollini, M.; Bartoli, F.; Marciano, A.; Zanca, R.; Slart, R.H.; Erba, P.A. Artificial intelligence and hybrid imaging: The best match for personalized medicine in oncology. *Eur. J. Hybrid Imaging* **2020**, *4*, 1–22. [[CrossRef](#)]
360. Nanda, V.; Dooley, S.; Singla, S.; Feizi, S.; Dickerson, J.P. Fairness through robustness: Investigating robustness disparity in deep learning. In Proceedings of the ACM Conference on Fairness, Accountability, and Transparency, Virtual Event, Canada, 3–10 March 2021; pp. 466–477.
361. Hamon, R.; Junklewitz, H.; Sanchez, I. Robustness and explainability of artificial intelligence. *Publ. Off. Eur. Union* **2020**, *207*, 2020.
362. Munoko, I.; Brown-Liburd, H.L.; Vasarhelyi, M. The ethical implications of using artificial intelligence in auditing. *J. Bus. Ethics* **2020**, *167*, 209–234. [[CrossRef](#)]
363. Adelakun, B.O. Ethical Considerations in the Use of AI for Auditing: Balancing Innovation and Integrity. *Eur. J. Account. Audit. Financ. Res.* **2022**, *10*, 91–108.
364. Mökander, J. Auditing of AI: Legal, ethical and technical approaches. *Digit. Soc.* **2023**, *2*, 49. [[CrossRef](#)]
365. Ashok, M.; Madan, R.; Joha, A.; Sivarajah, U. Ethical framework for Artificial Intelligence and Digital technologies. *Int. J. Inf. Manag.* **2022**, *62*, 102433. [[CrossRef](#)]
366. Xu, J. A review of self-supervised learning methods in the field of medical image analysis. *Int. J. Image Graph. Signal Process. (IJIGSP)* **2021**, *13*, 33–46. [[CrossRef](#)]
367. Taleb, A.; Lippert, C.; Klein, T.; Nabi, M. Multimodal self-supervised learning for medical image analysis. In Proceedings of the 27th International Conference on Information Processing in Medical Imaging, Virtual Event, 28–30 June 2021; pp. 661–673.

368. Zeebaree, S.R.; Ahmed, O.; Obid, K. Csaernet: An efficient deep learning architecture for image classification. In Proceedings of the 3rd International Conference on Engineering Technology and its Applications (IICETA), Najaf, Iraq, 6–7 September 2020; pp. 122–127.
369. Özyurt, F. Efficient deep feature selection for remote sensing image recognition with fused deep learning architectures. *J. Supercomput.* **2020**, *76*, 8413–8431. [[CrossRef](#)]
370. Jin, W.; Li, X.; Fatehi, M.; Hamarneh, G. Guidelines and evaluation of clinical explainable AI in medical image analysis. *Med. Image Anal.* **2023**, *84*, 102684. [[CrossRef](#)] [[PubMed](#)]
371. Han, S.H.; Kwon, M.S.; Choi, H.J. EXplainable AI (XAI) approach to image captioning. *J. Eng.* **2020**, *2020*, 589–594. [[CrossRef](#)]
372. Yang, G.; Rao, A.; Fernandez-Maloigne, C.; Calhoun, V.; Menegaz, G. Explainable AI (XAI) in biomedical signal and image processing: promises and challenges. In Proceedings of the IEEE International Conference on Image Processing (ICIP), Bordeaux, France, 16–19 October 2022; pp. 1531–1535.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.