# FPSNet: Focus-Perceptual-Semantic Full Flow Visual Redundancy Predicting for Camera Image

Xiongwei Xiao[1]✉

Independent Reseacher
xiongweixiaoxxw@gmail.com

**Abstract.** With the rapid popularization of electronic devices, the large amount of data generated by camera imaging poses a huge challenge to the limited storage capacity and communication bandwidth. Achieving higher compression ratios without sacrificing visual quality remains a fundamental challenge for image compression. In this paper, we propose a novel full flow bidirectional visual threshold estimation method for camera imaging perceptual compression. Specifically, we study the features from camera imaging to visual perception to semantic understanding, and characterize them with focus identification, perceptual distribution, and semantic segmentation respectively. We also carefully design feature extraction networks suitable for each feature type. In addition, we draw inspiration from the bidirectional perceptual mechanism of the human visual system and propose a feature extraction framework that adopts top-down and bottom-up methods. We further enhance our model by regulating and fusing bidirectional perceptual features through a gated decoding structure. Extensive experimental validation on benchmark datasets confirms that our FPSNet significantly improves the accuracy of visual redundancy prediction.

**Keywords:** Camera image · Visual redundancy · Image compression.

## 1 Instruction

In the digital era, camera imaging technology has become pervasive across various multimedia domains, including live streaming, security surveillance, and the film industry [1], [2], [3], [4]. These applications generate high-definition images, leading to substantial data production that challenges existing storage and transmission capacities. This issue is especially pertinent in real-time monitoring and mobile devices, where low latency and high data compression efficiency are imperative. Typically, traditional camera imaging involves the use of multiple optical sensors, which, while ensuring high-quality visual output, often introduce significant visual redundancy. Therefore, developing an efficient camera image compression method that minimizes redundant data without degrading image quality is essential for enhancing the performance of communication and storage systems. Visual Redundancy Prediction (VRP) technology addresses this

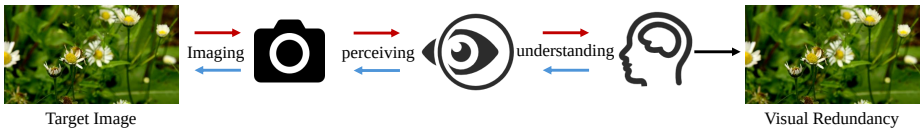


**Fig. 1.** The full flow of camera image perception

need by utilizing sophisticated image processing algorithms to accurately analyze and eliminate redundant information in camera imaging [7]. This approach not only improves the operational efficiency of multimedia applications but also establishes a new benchmark for real-time image compression in demanding environments. The full flow of camera images from imaging to perception, and then to understanding is shown in Figure 1.

Camera imaging remains a predominant method of image generation today. However, due to inherent equipment limitations and shooting conditions, such as focal length constraints and depth of field variations, different regions of camera images often exhibit varying degrees of focus, leading to differences in clarity [5]. Additionally, physiological characteristics of the human visual system (HVS), such as luminance adaptation and the concept of free energy [6], result in differing sensitivities to various image contents. Furthermore, the extent of human knowledge influences the level of understanding of targets within images, manifesting as varying degrees of attention to different content targets [12]. Therefore, a comprehensive consideration of these characteristics can significantly enhance the accuracy of visual redundancy prediction.

Existing VRP methods predominantly utilize the perceptual processes of the HVS for simulation. These methods are rooted in either handcrafted or deep learning paradigms, predicting visual redundancy through mathematical modeling or neural network fitting. However, accurately modeling the HVS's complex perceptual processes in VRP poses significant challenges, including representing various visual effects, understanding their interrelationships, and determining their integration mechanisms.

Both traditional handcrafted methods and deep learning approaches have their limitations in VRP: 1) Mathematical Modeling: Traditional handcrafted methods consider multiple visual effects related to perceptual redundancy, such as luminance adaptation [7], free energy [8], pattern complexity [6], visual field inhomogeneities [9], texture [10], and frequency domain [11]. However, the simple linear or nonlinear integration of these effects often compromises the accuracy of VRP predictions. 2) Deep Learning: These approachs employ the nonlinear fitting capabilities of deep neural networks to mimic the perceptual processes of the HVS and directly learns visual redundancy from images [12], [13], [14], [15], [16], [17], [18]. Further references include [19], [20], [21]. Despite their strengths, these methods may not fully utilize existing expert knowledge and often lack the analytical capabilities necessary for comprehensive image evaluation. Additionally, an exclusive focus on human visual perspectives may introduce biases in visual redundancy predictions.

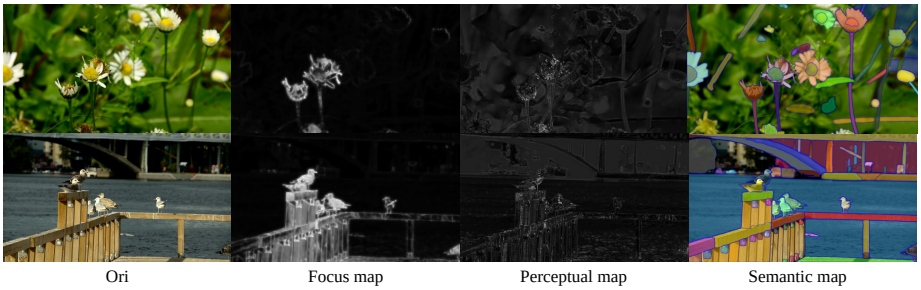


**Fig. 2.** Illustration of two camera images (first column), and their corresponding focus maps (second column), perceptual maps (third column), and semantic maps (fourth column). **Please zoom in for details**.

To address these issues, we propose a full-process camera imaging visual redundancy prediction method that integrates the advantages and limits the linearity of both approaches. This method is designed to fully utilize both existing visual perception expertise and the exceptional fitting capabilities of deep neural networks. It aims to consider the entire process of image creation, visual perception, and visual understanding, thereby fully comprehending the visual perceptual features of images across these three stages. The contributions of this work can be summarized as follows:

- To fully leverage the process from imaging to visual understanding, we have enriched the existing SHEN2020 [13] dataset. This enhancement involved processing the original images for defocus detection, analyzing visual perceptual characteristics, and conducting semantic segmentation to optimize the utilization of camera image features.
- To fully extract the features of camera images, we have designed specific network structures tailored to the distinct visual perception characteristics for the original image, the focus map, the perceptual map, and the semantic map, respectively.
- In alignment with the HVS's mechanisms, we designed a dual-stream bidirectional fusion network that integrates the well-established top-down(TD) and bottom-up(BU) visual perception processes. Additionally, we introduced a gating mechanism to regulate the integration of features from these dual streams, enhancing the network's efficacy in mimicking human visual understanding

## 2 Proposed Method

### 2.1 Overall framework

The overall network architecture of our proposed full-process visual redundancy prediction model for camera imaging is depicted in the accompanying figure.

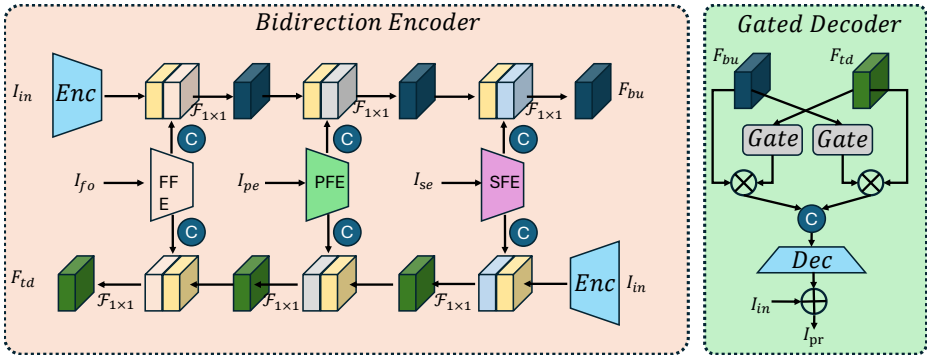


**Fig. 3.** Illustration of overall pipeline framework. The proposed FPSNet architecture including three meticulously designed feature extraction modules, a top-down and bottom-up dual-stream net and a gated decoder.

Initially, the architecture begins with an encoder section that features a bidirectional feature extraction and fusion network. This component is specifically designed to emulate the human visual understanding process effectively. Following this, we introduce three specialized feature encoders: the Focus Feature Encoder (FFE), the Perceptual Feature Encoder (PFE), and the Semantic Feature Encoder (SFE). These encoders are responsible for extracting features from the focus map $I_{fo}$, perceptual map $I_{pe}$, and semantic map $I_{se}$, respectively. The extracted features are integrated into the network in a sequence that adheres to both top-down and bottom-up processing paradigms. Conclusively, a gated network is implemented to regulate and synergize the bidirectional perceptual features, ensuring optimal feature integration and redundancy reduction. Our overall optimization objective, detailed in subsequent sections, aligns with these architectural frameworks to maximize efficiency and effectiveness in visual redundancy prediction.

$$\min \|\mathcal{F}_{FPS}(I_{in}, I_{fo}, I_{pe}, I_{se}) - I_{gt}\|_1, \tag{1}$$

where $I_{in}, I_{fo}, I_{pe}, I_{se}, I_{gt}$ denote input image, focus map, perceptual map, semantic map, and ground truth respectively. Next, we will detail our methodology module by module.

### 2.2 Feature Pre-extraction

**Focus Map:**Camera imaging is inherently constrained by the limitations of focal length and depth of field, which restrict the camera's ability to focus uniformly across various regions of a scene. For instance, when capturing the same scene with different focus settings, images can exhibit out-of-focus blur for objects that fall outside the depth of focus range, resulting in significantly different information content and visual perceptions. The areas within the focus range typically show enhanced clarity and contrast, whereas objects beyond this range

appear blurred due to defocus. Consequently, an accurate assessment of visual thresholds in camera imaging necessitates consideration of the focus distribution across the image. Inspired by these challenges, we propose a method to extract the focus distribution from camera images as a foundational step in constructing imaging features. As depicted in Figure 2, our focus distribution map is generated using the methodology described in [5]. This approach enables a more nuanced understanding of how focus affects image quality and visual perception, thereby improving the robustness and accuracy of our visual redundancy prediction model. Two examples are shown in the second column of Figure 2.

**Perceptual Map:**Distortion in various areas of an image leads to differing sensitivities in human perception, particularly in textured regions where the human eye may not consistently detect detail degradation, such as noise. Consequently, we have focused our research on quantifying the tolerance to visual changes to better understand these perceptual nuances. Due to factors such as luminance, color, and contrast, sensitivities to distortion vary across different areas of an image. Extensive research has established a significant correlation between the perception of distortion and image quality assessment. Hence, perceptible distortion serves as an indicator of the tolerance to quality changes in camera imaging images. To systematically measure the impact of quality perception differences in camera images, we propose the use of perceptible distortion as a metric. We adopt the method described in [6] to serve as our perceptual feature extractor. This approach allows us to effectively evaluate and quantify the perceptual impact of image quality variations. Two examples are shown in the third column of Figure 2.

**Semantic Map:**Different objects within an image exhibit distinct quality characteristics. Typically, humans are drawn to areas containing similar objects when viewing images. For instance, the perceived quality importance of persons and vehicles, which generally attract more attention, is considerably higher than that of less engaging elements like the sky or streets. To further understand these differences in quality perception across various semantic contents, we have integrated semantic information into our analytical framework. Research has demonstrated that semantic segmentation feature maps significantly enhance the effectiveness of visual redundancy prediction models [12]. Accordingly, we employ the BEIT model as our semantic feature extractor [22], enabling a more precise analysis of how semantic content influences perceived image quality. Two examples are shown in the forth column of Figure 2.

### 2.3 Network Design

**Focus Feature Encoder:**Focus areas in an image typically exhibit clearer textures, while out-of-focus areas present more blurred texture information, which increases with distance from the focus point. To effectively capture these focus distribution characteristics, we utilize the VGG network, a neural network architecture that has been proven effective in extracting texture features[23]. Moreover, since focus and out-of-focus areas often maintain specific spatial relationships, we have enhanced our approach by incorporating a spatial attention

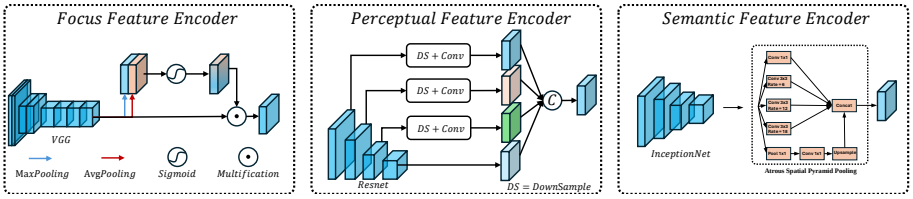


**Fig. 4.** Illustration of three feature extraction network structures: focus feture encoder (the left column), perceptual feature encoder (the middle column), semantic feature encoder (the right column).

module. This module is designed to modulate the focus distribution features, thereby refining the accuracy of our texture feature analysis and improving the overall performance of the system.

**Perceptual Feature Encoder:** HVS is sensitive to local distortions when the rest of the image exhibits fairly good quality [24], indicating that the quality of an image is determined by a combination of global and local features. Therefore, this prompts us to consider extracting both local and global visual perception features. Referring to the network structure in [24], we use features from different stages of ResNet [25] to characterize global and local perceptual features.

**Semantic Feature Encoder:**The theory of image pyramids demonstrates that human perception operates at various scales, indicating that humans experience differences in the level of detail they perceive in scenes and objects. For example, the amount of information presented by an object decreases as the camera distance increases, yet humans can still recognize the same object. Building on this discussion, we utilize Inception-Net [26], a neural network with multiple scales and receptive fields, to extract semantic features. Furthermore, to introduce a richer multi-scale feature set, we concatenate an Atrous Spatial Pyramid Pooling (ASPP) structure [27] at the end, effectively capturing multi-scale contextual information.

**Bidirection Encoder:** In the HVS, both TD and BU processing streams collectively influence how we perceive and interpret visual information. The BU process is sensory-driven, primarily involving the extraction of raw data and primary features from the visual scene, such as colors, edges, and shapes. These elements form the foundation of our visual perception, providing necessary information for higher-level interpretation. Conversely, the TD process is guided by our expectations, knowledge, and prior experiences, and it is used to interpret these basic features, constructing meaning and making decisions in complex scenes.

Based on the aforementioned theoretical foundation, we have adopted a bidirectional feature extraction network that aligns more closely with human visual perception processes while fully leveraging the fundamental features from focus, perceptual, and semantic. Specifically, we employ the same architecture used in the PFE as the feature encoder for the input image. The focus map, perceptual

map, and semantic map are sequentially integrated into the network. Each feature is then fused using a $1 \times 1$ convolution layer to ensure detailed preservation and integration of the diverse feature sets. Conversely, in the reverse process, the order of integration is inverted, but all other aspects of the fusion methodology remain unchanged. This symmetric approach allows for consistent feature processing while accommodating the different contributions of each feature type.

**Gated Decoder:** To more effectively harness the bidirectional features from both TD and BU processes, we have designed a gated fusion decoding structure. Specifically, this structure includes a gated perceptual fusion module that introduces a dynamic regulation mechanism. This mechanism allows high-level semantic information from the TD process to guide the integration of low-level features from the BU process. Consequently, the enriched low-level BU features enhance the comprehensiveness and depth of the TD understanding. This synergy optimizes the effectiveness and relevance of the fused features, leading to improved performance in feature-based tasks.

## 3 Experiment

### 3.1 Protocol

**Dataset Description:** In our experiments, we utilized the benchmark dataset SHEN2020 [13] which comprises 202 original images at a resolution of 1920x1080. Each image in this dataset includes one JND (Just Noticeable Difference) image, showcasing visually lossless compression distortions implemented via VVC encoding. A detailed description of the generation process for these images is available in [13]. The dataset was randomly divided into a training set and a test set at a 9:1 ratio. To further assess the generalization ability of various models, we also included four additional open-source benchmark datasets for cross-database validation: CSIQ [28], KADID10K [29], LIVE [30], and MCL-JCI [31].

**Comparison Methods:** To demonstrate the effectiveness of FPSNet, we conducted both qualitative and quantitative comparisons with seven prominent methods targeted at camera imaging: Yang2005SPIC [7], Liu2010TCSVT [10], Wu2013TMM [8], Wu2017TIP[6], Chen2019TCSVT [9], Shen2020TIP [13], and Jiang2022TIP [11]. We employed a noise injection method guided by the predicted visual redundancy distribution, randomly adding noise to each pixel as formalized by the following equation:

$$I_{dis} = I_{in} + \gamma \times I_{ra} \times I_{pr}, \tag{2}$$

where $I_{dis}$ represents the corrupted image, $I_{ra}$ represents a matrix of the same size as input image $I_{in}$, with each element randomly set to either -1 or 1, $I_{pr}$ represents the predicted visual redundancy map, and $\gamma$ represents an intensity adjustment scalar.

**Experimental Details:** Our FPSNet is implemented using the *PyTorch* deep learning framework. All weights are initialized using a truncated normal

**Table 1.** Comparison results of objective performance indicators SSIM for visual redundancy models, with the best result highlighted in **bold**.

| method | CSIQ | KADID-10K | LIVE | MCL-JCI | TestSet |
|---|---|---|---|---|---|
| Yang2005SPIC | 0.9515 | 0.9418 | 0.9439 | 0.9226 | 0.9449 |
| Liu2010SPIC | 0.9588 | 0.9443 | 0.9512 | 0.9289 | 0.9477 |
| Wu2013SPIC | 0.9489 | 0.9501 | 0.9565 | 0.9324 | 0.9453 |
| Wu2017SPIC | 0.9565 | 0.9528 | 0.9531 | 0.9369 | 0.9510 |
| Chen2019SPIC | 0.9611 | 0.9623 | 0.9614 | 0.9498 | 0.9612 |
| Shen2020SPIC | 0.9568 | 0.9453 | 0.9524 | 0.9369 | 0.9568 |
| Jiang2022SPIC | 0.9543 | 0.9657 | 0.9599 | **0.9564** | 0.9622 |
| Proposed | **0.9693** | **0.9744** | **0.9753** | 0.9559 | **0.9737** |

initializer. We employ the default parameters of the Adam optimizer, such as $\beta_1 = 0.9$ and $\beta_2 = 0.9$. Before training, we preprocess the input images by randomly cropping and rotating them to a size of 256x256. We set the batch size to 8 and the initial learning rate to 1e-5. The model training is conducted on an NVIDIA TITAN RTX GPU 24G GPU, which takes approximately 300 epochs to convergence. Feature preprocessing methods are extracted according to the open-source code provided in references[5, 24, 6].

### 3.2 Quantitative Analysis

Table 1 below presents the experimental results of our method compared to seven other prominent methods, evaluated under the SSIM/MS-SSIM quality assessment metrics with the distortion level set to MSE=100. As indicated by the data, our proposed method achieves competitive results across four datasets. These findings not only demonstrate our method's superior ability to preserve the structural information of images at equivalent noise levels but also highlight its effectiveness in maintaining image integrity. This represent the potential of our approach in applications demanding high fidelity and robust image analysis.

### 3.3 Qualitative Analysis

To more vividly illustrate the visual redundancy prediction performance of our method compared to competing approaches, the subsequent figure presents the visual redundancy values on test images, alongside distortion images after noise injection at MSE=100. In the visual redundancy distribution maps, areas of higher brightness indicate increased visual redundancy. The figure clearly shows that out-of-focus areas, regions with lower visual perceptual responses, and zones surrounding the target exhibit heightened visual redundancy. Conversely, regions that are more readily perceived by the human eye display less distortion, reflecting lower visual redundancy. Our method injects a more rational and effective distribution of visual redundancy across the images compared to other methods. This superior performance can be attributed to our method's comprehensive

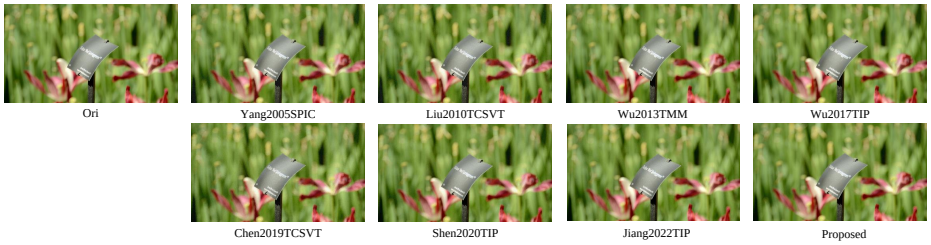


**Fig. 5.** Visual Quality Comparison of Contaminated Images Across Different Methods. **Please zoom in for details**.

**Table 2.** Ablation study of the three features, with the best result highlighted in **bold**.

| $I_{fo}$ | $I_{pe}$ | $I_{se}$ | PSNR | SSIM |
|---|---|---|---|---|
| ✗ | ✗ | ✗ | 34.8943 | 0.9602 |
| ✓ | ✗ | ✗ | 35.4238 | 0.9679 |
| ✗ | ✓ | ✗ | 35.5832 | 0.9719 |
| ✗ | ✗ | ✓ | 35.5828 | 0.9654 |
| ✓ | ✓ | ✓ | **35.6534** | **0.9737** |

integration of full-process features, from imaging to visual perception. Additionally, our approach adheres to both TD and BU human visual perception mechanisms and leverages the advanced nonlinear fitting capabilities of neural networks to more accurately emulate the complexities of the HVS.

### 3.4 Ablation Experiment

To further validate the robustness and effectiveness of our method, additional ablation experiments were conducted. FPSNet primarily incorporates three key features alongside the novel bidirectional network architecture and gated fusion module. Accordingly, ablation studies specifically targeted the input features and network modules to verify the necessity of the introduced features and the effectiveness of the network modules. We calculate the PSNR and SSIM between the groud truth and the predicted image to represent the accuracy of the network. The results of the ablation experiments are shown in the Table 2 and Table 3.

**Effects of three features:** To ensure the fairness of the experiments, all ablated input features were systematically replaced by the introduced features. The results reveal that the inclusion of perceptual features alone yields higher performance, likely attributable to the high relevance of these perceptual features to the HVS. When all three types of features are introduced, our method demonstrates superior performance, substantiating the necessity of incorporating all three full-process features.

**Effects of network:** For consistency in experimental parameters, two identical networks were used to emulate a unidirectional perceptual process, with the gating modulation module replaced by standard convolution layers. The findings

**Table 3.** Ablation study of the proposed nerwork, with the best result highlighted in **bold**.

| Bidirection Encoder | Gated Decoder | PSNR | SSIM |
|---|---|---|---|
| ✗ | ✗ | 34.5692 | 0.9603 |
| ✓ | ✗ | 35.4253 | 0.9676 |
| ✗ | ✓ | 35.5102 | 0.9711 |
| ✓ | ✓ | **35.6534** | **0.9737** |

confirm that the complete network architecture, including the bidirectional perceptual network and gated decoding module, exhibits the best performance. This not only validates the effectiveness of our introduced network components but also highlights their critical role in enhancing the overall system performance.

## 4 Conclusion

In this paper, we propose a full flow bidirectional perceptual redundancy prediction network for camera images, inspired by human visual perception mechanisms. We have rethink the paradigm of camera image redundancy prediction by considering the entire process from camera imaging and visual perception to semantic understanding. This comprehensive approach aims to effectively eliminate visual redundancy, fully leveraging the unique characteristics of camera images coupled with the intricacies of human visual perception and cognitive understanding. Additionally, we introduced a bidirectional perceptual framework and a gated modulation module to further emulate the sophisticated mechanisms of human visual perception. Experimental results demonstrate that our method surpasses seven other comparative methods in both qualitative and quantitative analyses on benchmark datasets. We believe that the proposed method can significantly enhance the quality of service in camera image applications such as live streaming, surveillance, and other real-time imaging fields.

## References

1. Barnett, Thomas, et al. "Cisco visual networking index (vni) complete forecast update, 2017–2022." Americas/EMEAR Cisco Knowledge Network (CKN) Presentation (2018): 1-30.
2. Huang, Siqi, Jiang Xie, and Muhana Magboul Ali Muslam. "A cloud computing based deep compression framework for UHD video delivery." IEEE Transactions on Cloud Computing 11.2 (2022): 1562-1574.
3. Menon, Vignesh V., et al. "CODA: Content-aware Frame Dropping Algorithm for High Frame-rate Video Streaming." DCC. 2022.
4. Zhao, Lei, et al. "Enhanced surveillance video compression with dual reference frames generation." IEEE Transactions on Circuits and Systems for Video Technology 32.3 (2021): 1592-1606.

5. Shi, Jianping, Li Xu, and Jiaya Jia. "Just noticeable defocus blur detection and estimation." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015.
6. Wu, Jinjian, et al. "Enhanced just noticeable difference model for images with pattern complexity." IEEE Transactions on Image Processing 26.6 (2017): 2682-2693.
7. Yang, Xiaokang, et al. "Just noticeable distortion model and its applications in video coding." Signal processing: Image communication 20.7 (2005): 662-680.
8. Wu, Jinjian, et al. "Just noticeable difference estimation for images with free-energy principle." IEEE Transactions on Multimedia 15.7 (2013): 1705-1710.
9. Chen, Zhenzhong, and Wei Wu. "Asymmetric foveated just-noticeable-difference model for images with visual field inhomogeneities." IEEE Transactions on Circuits and Systems for Video Technology 30.11 (2019): 4064-4074.
10. Liu, Anmin, et al. "Just noticeable difference for images with decomposition model for separating edge and textured regions." IEEE Transactions on Circuits and Systems for Video Technology 20.11 (2010): 1648-1652.
11. Jiang, Qiuping, et al. "Toward top-down just noticeable difference estimation of natural images." IEEE Transactions on Image Processing 31 (2022): 3697-3712.
12. Xie, Wuyuan, et al. "Just noticeable visual redundancy forecasting: a deep multimodal-driven approach." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 37. No. 3. 2023.
13. Shen, Xuelin, et al. "Just noticeable distortion profile inference: A patch-level structural visibility learning approach." IEEE Transactions on Image Processing 30 (2020): 26-38.
14. Liu, Huanhua, et al. "Deep learning-based picture-wise just noticeable distortion prediction model for image compression." IEEE Transactions on Image Processing 29 (2019): 641-656.
15. Zhang, Yun, et al. "Deep learning based just noticeable difference and perceptual quality prediction models for compressed video." IEEE Transactions on Circuits and Systems for Video Technology 32.3 (2021): 1197-1212.
16. Zhang, Xinfeng, et al. "Satisfied-user-ratio modeling for compressed video." IEEE Transactions on Image Processing 29 (2020): 3777-3789.
17. Nami, Sanaz, et al. "BL-JUNIPER: A CNN-assisted framework for perceptual video coding leveraging block-level JND." IEEE Transactions on Multimedia (2022).
18. Tian, Tao, et al. "Perceptual image compression with block-level just noticeable difference prediction." ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM) 16.4 (2021): 1-15.
19. Lin, Weisi, and Gheorghita Ghinea. "Progress and opportunities in modelling just-noticeable difference (JND) for multimedia." IEEE Transactions on Multimedia 24 (2021): 3706-3721.
20. Wang, Guoxiang, et al. "A survey on just noticeable distortion estimation and its applications in video coding." Journal of Visual Communication and Image Representation (2024): 104034.
21. Wu, Jinjian, Guangming Shi, and Weisi Lin. "Survey of visual just noticeable difference estimation." Frontiers of Computer Science 13 (2019): 4-15.
22. Bao, Hangbo, et al. "BEiT: BERT Pre-Training of Image Transformers." International Conference on Learning Representations. 2021.
23. Noh, Hyeonwoo, Seunghoon Hong, and Bohyung Han. "Learning deconvolution network for semantic segmentation." Proceedings of the IEEE international conference on computer vision. 2015.

24. Su, Shaolin, et al. "Blindly assess image quality in the wild guided by a self-adaptive hyper network." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.
25. He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
26. Szegedy, Christian, et al. "Inception-v4, inception-resnet and the impact of residual connections on learning." Proceedings of the AAAI conference on artificial intelligence. Vol. 31. No. 1. 2017.
27. Chen, Liang-Chieh, et al. "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs." IEEE transactions on pattern analysis and machine intelligence 40.4 (2017): 834-848.
28. Larson, Eric C., and Damon M. Chandler. "Most apparent distortion: full-reference image quality assessment and the role of strategy." Journal of electronic imaging 19.1 (2010): 011006-011006.
29. Lin, Hanhe, Vlad Hosu, and Dietmar Saupe. "KADID-10k: A large-scale artificially distorted IQA database." 2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX). IEEE, 2019.
30. Sheikh, H. 2005. LIVE image quality assessment database release 2. http://live. ece. utexas. edu/research/quality.
31. Wang, Haiqiang, et al. "MCL-JCV: a JND-based H. 264/AVC video quality assessment dataset." 2016 IEEE international conference on image processing (ICIP). IEEE, 2016.