Deep Learning and Computer Vision Papers

Luke Ottey
lottey98@gmail.com

Contents

1 ARCHITECTURES

1.1 first deep CNNs

(1)

1.2 building blocks

1.2.1 residual connections.

(1)

1.2.2 dense connections.

(1)

1.2.3 inception-style.

(1) Xception: Deep Learning with Depthwise Separable Convolutions [89]

1.2.4 drop-in and add-on components.

(1)

1.2.5 mobile-inverted bottleneck.

(1)

1.3 activations

(1) Searching for Activation Functions [373]

1.4 normalizations

1.4.1 weight normalization.

(1) Weight Normalization: A Simple Reparameterization to Accelerate Training of Deep Neural Networks [399]

1.5 other layers

(1) A ConvNet for the 2020s [295]

1.6 transformers

(1) Vision Transformers for Dense Prediction [374]
(2) LoFTR: Detector-Free Local Feature Matching with Transformers [442]

2 REGULARIZATION

2.1 augmentations

2.1.1 mixed-sampling.

(1)

2.1.2 automatic.

(1)

(1) A Simple Framework for Contrastive Learning of Visual Representations [77]
(2) What makes for good views for contrastive learning [460]

2.2 consistency

(1) A Simple Framework for Contrastive Learning of Visual Representations [77]

2.3 knowledge-distillation

(1) WSOD2: Learning Bottom-Up and Top-Down Objectness Distillation for Weakly-Supervised Object Detection [537]
(2) Comprehensive Attention Self-Distillation for Weakly-Supervised Object Detection [200]

2.4 intermediate-layer

(1)

(1) When Does Label Smoothing Help? [337]

3 ADVERSARIAL

3.1 generative adversarial networks

3.1.1 theory.

(1)

3.1.2 wasserstein.

(1)

3.1.3 maximum mean discrepancy.

(1)

3.1.4 conditional/auxililary.

(1)

3.1.5 image-to-image translation.

(1)

3.1.6 other popular variants.

(1)

(1) Robust and Accurate Object Detection via Adversarial Learning [82]

4 DETECTION

4.1 detectors

4.1.1 one-stage.

yolo.

(1) YOLACT: Real-Time Instance Segmentation [48]
(2) YOLOX: Exceeding YOLO Series in 2021 [136]
(3) You Only Look One-level Feature [76]
(4) PP-YOLOv2: A Practical Object Detector [198]

retinanet.

(1)

anchor-free.

(1)

4.1.2 *two-stage.*

rcnn.

(1)

4.1.3 *transformers.*

(1)

4.2 segmentation

4.2.1 *instance.*

(1) YOLACT: Real-Time Instance Segmentation [48]

4.2.2 *semantic.*

(1)

4.2.3 *panoptic.*

(1)

(1) Simple Does It: Weakly Supervised Instance and Semantic Segmentation [224]
(2) Weakly Supervised Learning of Instance Segmentation With Inter-Pixel Relations [4]

4.3 weakly-supervised learning

(1) WSOD2: Learning Bottom-Up and Top-Down Objectness Distillation for Weakly-Supervised Object Detection [537]
(2) Comprehensive Attention Self-Distillation for Weakly-Supervised Object Detection [200]

4.4 applied

(1) Depth CNNs for RGB-D Scene Recognition: Learning from Scratch Better than Transferring from RGB-CNNs [429]

4.5 anchor boxes

(1)

4.6 salient object detection

(1)

(1) Weakly Supervised Region Proposal Network and Object Detection [455]
(2) Robust and Accurate Object Detection via Adversarial Learning [82]

5 CONTRASTIVE LEARNING

(1) A Simple Framework for Contrastive Learning of Visual Representations [77]
(2) What makes for good views for contrastive learning [460]

6 SEMI-SUPERVISED LEARNING

6.1 pseudo-labeling

(1) Pseudo-Label : The Simple and Efficient Semi-Supervised Learning Method for Deep Neural Networks [255]

7 TRANSFER LEARNING

(1) What is being transferred in transfer learning? [343]

8 WEAKLY-SUPERVISED LEARNING

(1) Simple Does It: Weakly Supervised Instance and Semantic Segmentation [224]
(2) Weakly Supervised Learning of Instance Segmentation With Inter-Pixel Relations [4]
(3) Weakly Supervised Region Proposal Network and Object Detection [455]

9 X-SHOT LEARNING

9.1 few-shot learning

(1)

9.2 one-shot learning

(1)

9.3 zero-shot learning

(1) Zero-Shot Learning—A Comprehensive Evaluation of the Good, the Bad and the Ugly [515]

10 DIFFERENTIAL EVOLUTION

(1) Differential Evolution: A Survey of the State-of-the-Art [102]
(2) Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces [438]

11 LOW-LEVEL VISUAL PROCESSING

11.1 correspondence

(1)

11.2 feature matching

(1) LoFTR: Detector-Free Local Feature Matching with Transformers [442]

11.3 descriptors

(1)

12 STATISTICS

12.1 markov chains

12.1.1 monte-carlo.

(1)

(1) Principal component analysis [508]

13 NO-CATEGORY

(1) torch.fx: Practical Program Capture and Transformation for Deep Learning in Python [379]
(2) Verified Uncertainty Calibration [248]
(3) VOLO: Vision Outlooker for Visual Recognition [532]
(4) TensorMask: A Foundation for Dense Object Segmentation [81]
(5) The Lovasz-Softmax Loss: A Tractable Surrogate for the Optimization of the Intersection-Over-Union Measure in Neural Networks [36]
(6) The surprising impact of mask-head architecture on novel class segmentation [42]
(7) Transformers in Vision: A Survey [223]
(8) TrivialAugment: Tuning-free Yet State-of-the-Art Data Augmentation [338]
(9) Unbiased Teacher for Semi-Supervised Object Detection [294]
(10) Unsupervised Representation Learning by Predicting Image Rotations [141]
(11) Understanding Contrastive Representation Learning through Alignment and Uniformity on the Hypersphere [492]
(12) Understanding Data Augmentation for Classification: When to Warp? [509]
(13) Understanding image representations by measuring their equivariance and equivalence [259]
(14) Unsupervised Data Augmentation for Consistency Training [519]
(15) Unsupervised Feature Learning via Non-parametric Instance Discrimination [514]
(16) Unsupervised Object Segmentation by Redrawing [74]
(17) Unsupervised Visual Representation Learning by Context Prediction [109]
(18) Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles [252]
(19) Spatial Transformer Networks [207]
(20) SpineNet: Learning Scale-Permuted Backbone for Recognition and Localization [115]
(21) Structured Pruning of Deep Convolutional Neural Networks [12]
(22) Supervised Contrastive Learning [225]
(23) S4L: Self-Supervised Semi-Supervised Learning [540]
(24) ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design [306]
(25) Side-Aware Boundary Localization for More Precise Object Detection [490]
(26) Simple Copy-Paste is a Strong Data Augmentation Method for Instance Segmentation [138]
(27) Self-Normalizing Neural Networks [234]
(28) Self-Supervised Learning of Pretext-Invariant Representations [323]
(29) Self-Supervised Equivariant Attention Mechanism for Weakly Supervised Semantic Segmentation [499]
(30) Look-Into-Object: Self-Supervised Structure Modeling for Object Recognition [567]
(31) Semantic Instance Segmentation with a Discriminative Loss Function [53]
(32) SPICE: Semantic Pseudo-labeling for Image Clustering [347]
(33) Rethinking Channel Dimensions for Efficient Model Design [160]
(34) Robust and Accurate Object Detection via Adversarial Learning [82]
(35) Sharpness-Aware Minimization for Efficiently Improving Generalization [129]
(36) Scale-Aware Trident Networks for Object Detection [271]
(37) Selective Kernel Networks [269]
(38) Scaled-YOLOv4: Scaling Cross Stage Partial Network [483]
(39) Rethinking Pre-training and Self-training [577]
(40) Rethinking "Batch" in BatchNorm [513]

(41) Revisiting ResNets: Improved Training and Scaling Strategies [32]
(42) Region Proposal by Guided Anchoring [487]
(43) Representation Learning with Contrastive Predictive Coding [472]
(44) ResNeSt: Split-Attention Networks [545]
(45) Research on a Surface Defect Detection Algorithm Based on MobileNet-SSD [272]
(46) Probabilistic Anchor Assignment with IoU Prediction for Object Detection [212]
(47) CenterMask: Real-Time Anchor-Free Instance Segmentation [256]
(48) Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network [412]
(49) Receptive Field Block Net for Accurate and Fast Object Detection [290]
(50) Provably Efficient Online Hyperparameter Optimization with Population-Based Bandits [358]
(51) Pseudo-mask Matters in Weakly-supervised Semantic Segmentation [273]
(52) ReMixMatch: Semi-Supervised Learning with Distribution Matching and Augmentation Anchoring [37]
(53) Rectifying Pseudo Label Learning via Uncertainty Estimation for Domain Adaptive Semantic Segmentation [564]
(54) Pseudo-Labeling and Confirmation Bias in Deep Semi-Supervised Learning [13]
(55) PatchShuffle Regularization [215]
(56) Pretrained Transformers Improve Out-of-Distribution Robustness [179]
(57) PAG-YOLO: A Portable Attention-Guided YOLO Network for Small Ship Detection [190]
(58) Path Aggregation Network for Instance Segmentation [291]
(59) Momentum Contrast for Unsupervised Visual Representation Learning [165]
(60) Moser Flow: Divergence-based Generative Modeling on Manifolds [393]
(61) Multi-Similarity Loss With General Pair Weighting for Deep Metric Learning [495]
(62) NAS-FPN: Learning Scalable Feature Pyramid Architecture for Object Detection [140]
(63) Network size and weights size for memorization with two-layers neural networks [59]
(64) Self-Training With Noisy Student Improves ImageNet Classification [520]
(65) No Fuss Distance Metric Learning Using Proxies [331]
(66) Non-Local Neural Networks With Grouped Bilinear Attentional Transforms [85]
(67) Non-local Neural Networks [494]
(68) On the Relationship between Self-Attention and Convolutional Layers [97]
(69) Object Region Mining with Adversarial Erasing: A Simple Classification to Semantic Segmentation Approach [503]
(70) Objects as Points [568]
(71) On Calibration of Modern Neural Networks [157]
(72) On Mixup Training: Improved Calibration and Predictive Uncertainty for Deep Neural Networks [459]
(73) On Network Design Spaces for Visual Recognition [370]
(74) On the Expressive Power of Deep Learning: A Tensor Analysis [94]
(75) OCGAN: One-Class Novelty Detection Using GANs With Constrained Latent Representations [362]
(76) Optimal Adaptive and Accelerated Stochastic Gradient Descent [105]
(77) Optuna: A Next-generation Hyperparameter Optimization Framework [6]
(78) Modeling Visual Context is Key to Augmenting Object Detection Datasets [119]
(79) MLP-Mixer: An all-MLP Architecture for Vision [464]
(80) MMDetection: Open MMLab Detection Toolbox and Benchmark [71]
(81) MobileViT: Light-weight, General-purpose, and Mobile-friendly Vision Transformer [315]
(82) Making Convolutional Networks Shift-Invariant Again [550]
(83) Meta Pseudo Labels [365]

(84) MetaAnchor: Learning to Detect Objects with Customized Anchors [525]
(85) MixConv: Mixed Depthwise Convolutional Kernels [451]
(86) MixUp as Locally Linear Out-Of-Manifold Regularization [158]
(87) Learning Data Augmentation Strategies for Object Detection [576]
(88) Learning Depth from Single Monocular Images Using Deep Convolutional Neural Fields [286]
(89) Learning Efficient Object Detection Models with Knowledge Distillation [70]
(90) Learning Representations by Maximizing Mutual Information Across Views [22]
(91) Learning Transferable Architectures for Scalable Image Recognition [579]
(92) Learning to Segment Every Thing [191]
(93) Learning to Segment Object Candidates [366]
(94) Learning to Segment via Cut-and-Paste [382]
(95) Least Squares Generative Adversarial Networks [312]
(96) Towards Automated Deep Learning: Efficient Joint Neural Architecture and Hyperparameter Search [536]
(97) Large Batch Training of Convolutional Networks [528]
(98) Do Better ImageNet Models Transfer Better? [239]
(99) LambdaNetworks: Modeling Long-Range Interactions Without Attention [31]
(100) Invariant Information Clustering for Unsupervised Image Classification and Segmentation [211]
(101) Instance-Aware, Context-Focused, and Memory-Efficient Weakly Supervised Object Detection [385]
(102) Improvements to Context Based Self-Supervised Learning [339]
(103) Improved Deep Metric Learning with Multi-class N-pair Loss Objective [426]
(104) Soft-NMS — Improving Object Detection with One Line of Code [47]
(105) Improving Robustness Without Sacrificing Accuracy with Patch Gaussian Augmentation [299]
(106) Improving the Robustness of Deep Neural Networks via Stability Training [562]
(107) InstaBoost: Boosting Instance Segmentation via Probability Map Guided Copy-Pasting [126]
(108) BoxInst: High-Performance Instance Segmentation with Box Annotations [463]
(109) High-Performance Large-Scale Image Recognition Without Normalization [57]
(110) Imbalance Problems in Object Detection: A Review [353]
(111) GhostNet: More Features From Cheap Operations [162]
(112) Gradient Starvation: A Learning Proclivity in Neural Networks [363]
(113) Group Invariance, Stability to Deformations, and Complexity of Deep Convolutional Representations [41]
(114) FreeAnchor: Learning to Match Anchors for Visual Object Detection [553]
(115) FoveaBox: Beyound Anchor-Based Object Detection [238]
(116) Gather-Excite: Exploiting Feature Context in Convolutional Neural Networks [188]
(117) Generalisation in humans and deep neural networks [137]
(118) Feature Pyramid Networks for Object Detection [283]
(119) Filter Response Normalization Layer: Eliminating Batch Dependence in the Training of Deep Neural Networks [420]
(120) Fast R-CNN [144]
(121) FCOS: Fully Convolutional One-Stage Object Detection [461]
(122) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks [384]
(123) Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression [563]
(124) Fast and Accurate Model Scaling [110]
(125) Dynamic Group Convolution for Accelerating Convolutional Neural Networks [440]
(126) Ensemble Adversarial Training: Attacks and Defenses [467]
(127) Explainable Deep One-Class Classification [296]
(128) Efficient Neural Architecture Search via Parameter Sharing [364]

(129) EfficientDet: Scalable and Efficient Object Detection [452]
(130) Emerging Properties in Self-Supervised Vision Transformers [67]
(131) Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation [73]
(132) End-to-End Object Detection with Transformers [65]
(133) Evolving Normalization-Activation Layers [287]
(134) FaceNet: A unified embedding for face recognition and clustering [405]
(135) Densely Connected Convolutional Networks [193]
(136) Designing Network Design Spaces [371]
(137) On the Performance of Differential Evolution for Hyperparameter Tuning [404]
(138) Deep One-Class Classification [395]
(139) Deep learning is adaptive to intrinsic dimensionality of model smoothness in anisotropic Besov space [444]
(140) Deep Metric Learning With Tuplet Margin Loss [529]
(141) Deep Networks with Stochastic Depth [194]
(142) CBAM: Convolutional Block Attention Module [510]
(143) CSPNet: A New Backbone that can Enhance Learning Capability of CNN [485]
(144) Channel Pruning for Accelerating Very Deep Neural Networks [173]
(145) Characterizing signal propagation to close the performance gap in unnormalized ResNets [56]
(146) Classification is a Strong Baseline for Deep Metric Learning [538]
(147) Compact Bilinear Pooling [133]
(148) Concurrent Spatial and Channel Squeeze & Excitation in Fully Convolutional Networks [392]
(149) Convolutional Rectifier Networks as Generalized Tensor Decompositions [95]
(150) CornerNet: Detecting Objects as Paired Keypoints [253]
(151) CIAN: Cross-Image Affinity Net for Weakly Supervised Semantic Segmentation [125]
(152) Cyclical Learning Rates for Training Neural Networks [421]
(153) D2Det: Towards High Quality Object Detection and Instance Segmentation [64]
(154) Deep Anomaly Detection with Outlier Exposure [180]
(155) Decoupled Weight Decay Regularization [301]
(156) Deformable DETR: Deformable Transformers for End-to-End Object Detection [574]
(157) Data-Efficient Image Recognition with Contrastive Predictive Coding [175]
(158) Deep Clustering for Unsupervised Learning of Visual Features [66]
(159) Deep Feature Pyramid Reconfiguration for Object Detection [237]
(160) CCNet: Criss-Cross Attention for Semantic Segmentation [199]
(161) Consistency Regularization for Generative Adversarial Networks [547]
(162) AutoAssign: Differentiable Label Assignment for Dense Object Detection [571]
(163) Bridging the Gap Between Anchor-Based and Anchor-Free Detection via Adaptive Training Sample Selection [551]
(164) Averaging Weights Leads to Wider Optima and Better Generalization [204]
(165) Benchmarking Neural Network Robustness to Common Corruptions and Perturbations [176]
(166) Bag of Freebies for Training Object Detection Neural Networks [558]
(167) Bilinear CNNs for Fine-grained Visual Recognition [285]
(168) Billion-scale semi-supervised learning for image classification [523]
(169) Bottleneck Transformers for Visual Recognition [434]
(170) Bounding boxes for weakly supervised segmentation: Global constraints get close to full supervision [221]
(171) Bag of Tricks for Image Classification with Convolutional Neural Networks [171]
(172) Best practices for convolutional neural networks applied to visual document analysis [415]
(173) Big Self-Supervised Models are Strong Semi-Supervised Learners [78]

(174) Attention-Based Dropout Layer for Weakly Supervised Single Object Localization and Semantic Segmentation [87]
(175) Attention-guided Context Feature Pyramid Network for Object Detection [63]
(176) Augmentation for small object detection [233]
(177) Adversarial Complementary Learning for Weakly Supervised Object Localization [554]
(178) Adversarial Examples Improve Image Recognition [517]
(179) Adversarial Logit Pairing [216]
(180) Adversarial Training Can Hurt Generalization [372]
(181) Adversarial Training for Free! [409]
(182) An Energy and GPU-Computation Efficient Backbone Network for Real-Time Object Detection [257]
(183) Anchor Box Optimization for Object Detection [566]
(184) Anchor DETR: Query Design for Transformer-Based Detector [500]
(185) A survey on Image Data Augmentation for Deep Learning [413]
(186) A survey on addressing high-class imbalance in big data [258]
(187) A systematic study of the class imbalance problem in convolutional neural networks [61]
(188) An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale [112]
(189) A generalized framework for population based training [263]
(190) Revisiting dilated convolution: A simple approach for weakly-and semi-supervised semantic segmentation [504]
(191) Calibrating deep neural networks using focal loss [335]
(192) A survey on semi-supervised learning [473]
(193) Consistency regularization for generative adversarial networks [546]
(194) Large scale GAN training for high fidelity natural image synthesis [58]
(195) Temporal ensembling for semi-supervised learning [251]
(196) Training convolutional networks with noisy labels [441]
(197) What is being transferred in transfer learning? [344]
(198) Pretrained transformers improve out-of-distribution robustness [178]
(199) Pelee: A real-time object detection system on mobile devices [491]
(200) Differential evolution: A recent review based on state-of-the-art works [3]
(201) Efficient inference in fully connected crfs with gaussian edge potentials [242]
(202) Efficientnetv2: Smaller models and faster training [449]
(203) A critical analysis of self-supervision, or what we can learn from a single image [19]
(204) A combinatorial perspective on transfer learning [488]
(205) A large-scale study of representation learning with the visual task adaptation benchmark [541]
(206) Bayesian design of control space for optimal assimilation of observations. Part I: Consistent multiscale formalism [46]
(207) A system for massively parallel hyperparameter tuning [267]
(208) Realistic evaluation of deep semi-supervised learning algorithms [355]
(209) Virtual adversarial training: a regularization method for supervised and semi-supervised learning [326]
(210) Deep clustering with a dynamic autoencoder: From reconstruction towards centroids construction [332]
(211) Fixmatch: Simplifying semi-supervised learning with consistency and confidence [427]
(212) Mixmatch: A holistic approach to semi-supervised learning [38]
(213) Deep neural nets with interpolating function as output activation [481]
(214) There are many consistent explanations of unlabeled data: Why you should average [20]
(215) Interpolation consistency training for semi-supervised learning [476]
(216) Charting the right manifold: Manifold mixup for few-shot learning [310]

(217) Un-mix: Rethinking image mixtures for unsupervised visual representation learning [411]
(218) Deformable convnets v2: More deformable, better results [573]
(219) CSPNet: A new backbone that can enhance learning capability of CNN [484]
(220) Understanding image representations by measuring their equivariance and equivalence [260]
(221) Group normalization [512]
(222) Yolov4: Optimal speed and accuracy of object detection [45]
(223) Generalized intersection over union: A metric and a loss for bounding box regression [387]
(224) Clustergan: Latent space clustering in generative adversarial networks [334]
(225) NCP-VAE: Variational autoencoders with noise contrastive priors [11]
(226) Invariance and stability of deep convolutional representations [40]
(227) The large learning rate phase of deep learning: the catapult mechanism [262]
(228) Detnet: A backbone network for object detection [277]
(229) Three factors influencing minima in sgd [208]
(230) A disciplined approach to neural network hyper-parameters: Part 1–learning rate, batch size, momentum, and weight decay [423]
(231) Mish: A self regularized non-monotonic neural activation function [322]
(232) Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs) [93]
(233) Deep Sparse Rectifier Neural Networks [146]
(234) Generalization in deep learning [219]
(235) mixup: Beyond empirical risk minimization [543]
(236) Manifold mixup: Better representations by interpolating hidden states [477]
(237) Cutmix: Regularization strategy to train strong classifiers with localizable features [534]
(238) Augmix: A simple data processing method to improve robustness and uncertainty [181]
(239) Improved regularization of convolutional neural networks with cutout [107]
(240) Dropout: a simple way to prevent neural networks from overfitting [436]
(241) Dropblock: A regularization method for convolutional networks [139]
(242) Shakedrop regularization for deep residual learning [524]
(243) Shake-shake regularization [135]
(244) Autoaugment: Learning augmentation policies from data [98]
(245) Fast autoaugment [281]
(246) Circumventing Outliers of AutoAugment with Knowledge Distillation [502]
(247) Population based augmentation: Efficient learning of augmentation policy schedules [183]
(248) Randaugment: Practical automated data augmentation with a reduced search space [99]
(249) Spatial transformer networks [206]
(250) Fully convolutional networks for semantic segmentation [297]
(251) U-net: Convolutional networks for biomedical image segmentation [391]
(252) CSPNet: A new backbone that can enhance learning capability of cnn [484]
(253) Complex-yolo: An euler-region-proposal for real-time 3d object detection on point clouds [416]
(254) Deep learning for generic object detection: A survey [288]
(255) Efficientdet: Scalable and efficient object detection [453]
(256) Faster r-cnn: Towards real-time object detection with region proposal networks [383]
(257) Fast r-cnn [142]
(258) Mask r-cnn [166]
(259) Rich feature hierarchies for accurate object detection and semantic segmentation [143]
(260) Fcos: Fully convolutional one-stage object detection [462]
(261) Feature pyramid networks for object detection [282]

(262) Focal loss for dense object detection [284]
(263) Objects as points [569]
(264) Path aggregation network for instance segmentation [292]
(265) Receptive field block net for accurate and fast object detection [289]
(266) Speed/accuracy trade-offs for modern convolutional object detectors [196]
(267) Ssd: Single shot multibox detector [293]
(268) Towards multi-class object detection in unconstrained remote sensing imagery [21]
(269) Unsupervised co-segmentation through region matching [394]
(270) Unsupervised object discovery and localization in the wild: Part-based matching with bottom-up region proposals [86]
(271) You only look once: Unified, real-time object detection [376]
(272) YOLO9000: better, faster, stronger [377]
(273) Yolov3: An incremental improvement [378]
(274) YOLOv4: Optimal Speed and Accuracy of Object Detection [45]
(275) Adaptis: Adaptive instance selection network [425]
(276) Panoptic segmentation [232]
(277) Weakly-and semi-supervised panoptic segmentation [268]
(278) Upsnet: A unified panoptic segmentation network [522]
(279) Attention-guided unified network for panoptic segmentation [270]
(280) Panoptic feature pyramid networks [232]
(281) PP-YOLO: An Effective and Efficient Implementation of Object Detector [298]
(282) SOLOv2: Dynamic, Faster and Stronger [497]
(283) Solo: Segmenting objects by locations [496]
(284) Bayesian segnet: Model uncertainty in deep convolutional encoder-decoder architectures for scene understanding [220]
(285) Calibrated Adversarial Refinement for Multimodal Semantic Segmentation [218]
(286) DCNAS: Densely Connected Neural Architecture Search for Semantic Image Segmentation [555]
(287) Object-contextual representations for semantic segmentation [533]
(288) Rethinking atrous convolution for semantic image segmentation [72]
(289) The role of context for object detection and semantic segmentation in the wild [330]
(290) An uncertain future: Forecasting from static images using variational autoencoders [480]
(291) Auto-encoding variational bayes [230]
(292) Do better imagenet models transfer better? [240]
(293) How transferable are features in deep neural networks? [527]
(294) Large scale fine-grained categorization and domain-specific transfer learning [100]
(295) Syn2real: A new benchmark forsynthetic-to-real visual domain adaptation [360]
(296) A theory of learning from different domains [33]
(297) Adversarial discriminative domain adaptation [469]
(298) Analysis of representations for domain adaptation [34]
(299) Deep domain confusion: Maximizing for domain invariance [470]
(300) Deep visual domain adaptation: A survey [481]
(301) Discriminative Feature Alignment: Improving Transferability of Unsupervised Domain Adaptation by Gaussian-guided Latent Alignment [486]
(302) Dlid: Deep learning for domain adaptation by interpolating between domains [90]
(303) Domain adaptive faster r-cnn for object detection in the wild [83]
(304) Domain separation networks [52]

(305) Domain-adversarial neural networks [5]
(306) Domain-adversarial training of neural networks [132]
(307) Marginalized denoising autoencoders for domain adaptation [75]
(308) Multi-adversarial faster-rcnn for unrestricted object detection [174]
(309) Strong-weak distribution alignment for adaptive object detection [397]
(310) Unsupervised domain adaptation by backpropagation [131]
(311) Argoverse: 3D Tracking and Forecasting with Rich Maps [68]
(312) How much real data do we actually need: Analyzing object detection performance using synthetic and real data [350]
(313) nuScenes: A multimodal dataset for autonomous driving [62]
(314) Xception: Deep Learning with Depthwise Separable Convolutions [88]
(315) Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition [167]
(316) Pyramidal Convolution: Rethinking Convolutional Neural Networks for Visual Recognition [118]
(317) Rethinking the Inception Architecture for Computer Vision [447]
(318) Benchmark Analysis of Representative Deep Neural Network Architectures [39]
(319) Deep Layer Aggregation [530]
(320) Understanding Deep Architectures using a Recursive Convolutional Network [122]
(321) Learning Implicitly Recurrent CNNs Through Parameter Sharing [401]
(322) Group Equivariant Convolutional Networks [96]
(323) Wide Residual Networks [535]
(324) Striving for Simplicity: The All Convolutional Net [432]
(325) FitNets: Hints for Thin Deep Nets [390]
(326) CBAM: Convolutional Block Attention Module [511]
(327) Squeeze-and-Excitation Networks [189]
(328) PolyNet: A Pursuit of Structural Diversity in Very Deep Networks [552]
(329) Maxout Networks [153]
(330) Highway Networks [437]
(331) Doubly Convolutional Neural Networks [539]
(332) Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning [445]
(333) Going Deeper with Convolutions [446]
(334) Imagenet classification with deep convolutional neural networks [247]
(335) Deep Pyramidal Residual Networks [159]
(336) MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications [187]
(337) ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices [556]
(338) Very Deep Convolutional Networks for Large-Scale Image Recognition [417]
(339) Aggregated Residual Transformations for Deep Neural Networks [521]
(340) Deep Residual Learning for Image Recognition [168]
(341) Densely Connected Convolutional Networks [192]
(342) Identity Mappings in Deep Residual Networks [170]
(343) EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks [450]
(344) Dense Semantic Correspondence where Every Pixel is a Classifier [55]
(345) FCSS: Fully Convolutional Self-Similarity for Dense Semantic Correspondence [228]
(346) Image Patch Matching Using Convolutional Descriptors with Euclidean Distance [316]
(347) Working hard to know your neighbor's margins: Local descriptor learning loss [320]
(348) PARN: Pyramidal Affine Regression Networks for Dense Semantic Correspondence [210]
(349) Universal Correspondence Network [91]

(350) Colour and illumination in computer vision [128]
(351) Learning multiple layers of features from tiny images [246]
(352) Multi-digit number recognition from street view imagery using deep convolutional neural networks [152]
(353) Harmonic Networks with Limited Training Samples [471]
(354) Principles of neural science [214]
(355) Object Recognition with and without Objects [575]
(356) Deep networks with stochastic depth [195]
(357) A downsampled variant of imagenet as an alternative to the cifar datasets [92]
(358) The caltech-ucsd birds-200-2011 dataset [479]
(359) 3d object representations for fine-grained categorization [244]
(360) Fine-grained visual classification of aircraft [309]
(361) Imagenet large scale visual recognition challenge [396]
(362) Beyond Filters: Compact Feature Map for Portable Deep Model [498]
(363) Compression-aware Training of Deep Networks [9]
(364) GhostNet: More Features from Cheap Operations [161]
(365) Be Your Own Teacher: Improve the Performance of Convolutional Neural Networks via Self Distillation [549]
(366) PerforatedCNNs: Acceleration through Elimination of Redundant Convolutions [127]
(367) Soft-to-Hard Vector Quantization for End-to-End Learning Compressible Representations [2]
(368) Wasserstein Distance Guided Representation Learning for Domain Adaptation [410]
(369) A Note on the Inception Score [26]
(370) Adversarial Feature Learning [111]
(371) cGANs with Projection Discriminator [325]
(372) Conditional Generative Adversarial Nets [319]
(373) Conditional Image Synthesis With Auxiliary Classifier GANs [352]
(374) Context Encoders: Feature Learning by Inpainting [359]
(375) Deep Generative Image Models using a Laplacian Pyramid of Adversarial Networks [106]
(376) Demystifying MMD GANs [43]
(377) Differentiable Augmentation for Data-Efficient GAN Training [560]
(378) DualGAN: Unsupervised Dual Learning for Image-to-Image Translation [526]
(379) Energy-based Generative Adversarial Network [559]
(380) f-GAN: Training Generative Neural Samplers using Variational Divergence Minimization [349]
(381) From GAN to WGAN [505]
(382) GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium [182]
(383) Generalization and Equilibrium in Generative Adversarial Nets (GANs) [18]
(384) InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets [80]
(385) SeqGAN: Sequence Generative Adversarial Nets with Policy Gradient [531]
(386) Activation Maximization Generative Adversarial Nets [570]
(387) Triple Generative Adversarial Nets [264]
(388) Continual Learning in Generative Adversarial Nets [406]
(389) Dual Discriminator Generative Adversarial Nets [346]
(390) Generative Adversarial Networks in Computer Vision: A Survey and Taxonomy [501]
(391) Generative Adversarial Text to Image Synthesis [380]
(392) Generative Modeling by Estimating Gradients of the Data Distribution [430]
(393) Generative Moment Matching Networks [274]

(394) Image-to-Image Translation with Conditional Adversarial Networks [203]
(395) Improved Techniques for Training GANs [398]
(396) Improved Training of Wasserstein GANs [156]
(397) Large Scale GAN Training for High Fidelity Natural Image Synthesis [58]
(398) Learning in Implicit Generative Models [327]
(399) Least Squares Generative Adversarial Networks [311]
(400) Loss-Sensitive Generative Adversarial Networks on Lipschitz Densities [368]
(401) MMD GAN: Towards Deeper Understanding of Moment Matching Network [265]
(402) Progressive Growing of GANs for Improved Quality, Stability, and Variation [217]
(403) Self-Attention Generative Adversarial Networks [544]
(404) Self-Supervised GANs via Auxiliary Rotation Loss [79]
(405) Semantic Image Synthesis with Spatially-Adaptive Normalization [357]
(406) Semi-Supervised Learning with Generative Adversarial Networks [351]
(407) Spectral Normalization for Generative Adversarial Networks [324]
(408) Twin Auxiliary Classifiers GAN [148]
(409) Towards Principled Methods for Training Generative Adversarial Networks [15]
(410) Training generative neural networks via Maximum Mean Discrepancy optimization [120]
(411) Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks [572]
(412) Unrolled Generative Adversarial Networks [317]
(413) Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks [369]
(414) Wasserstein GAN [16]
(415) Sphere Generative Adversarial Network Based on Geometric Moment Matching [356]
(416) On Convergence and Stability of GANs [235]
(417) Constraint-Aware Deep Neural Network Compression [69]
(418) Distinctive Image Features from Scale-Invariant Keypoints [303]
(419) SURF: Speeded Up Robust Features [27]
(420) On Detection of Multiple Object Instances Using Hough Transforms [25]
(421) Key.Net: Keypoint Detection by Handcrafted and Learned CNN Filters [250]
(422) KAZE Features [7]
(423) Beyond Cartesian Representations for Local Descriptors [121]
(424) Anatomy of the SIFT Method [386]
(425) Saliency filters: Contrast based filtering for salient region detection [361]
(426) Salient object detection: A survey [49]
(427) Salient Object Detection in the Deep Learning Era: An In-Depth Survey [493]
(428) RC-DARTS: Resource Constrained Differentiable Architecture Search [213]
(429) Neural Architecture Search with Reinforcement Learning [578]
(430) Joint Neural Architecture Search and Quantization [84]
(431) Designing Neural Network Architectures using Reinforcement Learning [23]
(432) AutoGAN: Neural Architecture Search for Generative Adversarial Networks [149]
(433) Binding in short-term visual memory. [506]
(434) A Model for Visual Memory Tasks1 [431]
(435) Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex [164]
(436) Integration of Local Features into Global Shapes Monkey and Human fMRI Studies [241]
(437) Matching Categorical Object Representations in Inferior Temporal Cortex of Man and Monkey [245]
(438) Object recognition with features inspired by visual cortex [408]
(439) Perceptual learning depends on perceptual constancy [134]

(440) Robust Object Recognition with Cortex-Like Mechanisms [407]
(441) Scene Perception in the Human Brain. [124]
(442) Visual Object Representation: Interpreting Neurophysiological Data within a Computational Framework [367]
(443) A Differential Equation for Modeling Nesterov's Accelerated Gradient Method: Theory and Insights [439]
(444) A disciplined approach to neural network hyper-parameters: Part 1 - learning rate, batch size, momentum, and weight decay [422]
(445) A geometric alternative to Nesterov's accelerated gradient descent [60]
(446) A Kronecker-factored approximate Fisher matrix for convolution layers [155]
(447) Adam: A Method for Stochastic Optimization [229]
(448) Gradient Descent Provably Optimizes Over-parameterized Neural Networks [114]
(449) Incorporating Nesterov Momentum into Adam [113]
(450) Newton Methods for Convolutional Neural Networks [482]
(451) Second-order Optimization for Neural Networks [313]
(452) SGDR: Stochastic Gradient Descent with Warm Restarts [300]
(453) Stochastic Gradient Descent Tricks [50]
(454) Three Factors Influencing Minima in SGD [209]
(455) Train longer, generalize better: closing the generalization gap in large batch training of neural networks [184]
(456) Understanding the difficulty of training deep feedforward neural networks [145]
(457) Variants of RMSProp and Adagrad with Logarithmic Regret Bounds [336]
(458) Bridging the Gaps Between Residual Learning, Recurrent Neural Networks and Visual Cortex [280]
(459) A Baseline for Detecting Misclassified and Out-of-Distribution Examples in Neural Networks [177]
(460) A guide to convolution arithmetic for deep learning [117]
(461) All you need is a good init [321]
(462) Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift [202]
(463) Convergence Analysis of Two-layer Neural Networks with ReLU Activation [276]
(464) Convolutional Neural Fabrics [403]
(465) Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification [169]
(466) Do we need hundreds of classifiers to solve real world classification problems? [104]
(467) Don't Decay the Learning Rate, Increase the Batch Size [424]
(468) Enhancing The Reliability of Out-of-distribution Image Detection in Neural Networks [278]
(469) Explaining nonlinear classification decisions with deep Taylor decomposition [328]
(470) Fourier Features Let Networks Learn High Frequency Functions in Low Dimensional Domains [454]
(471) Generalized Linear Models [314]
(472) Maximum Mean Discrepancy Gradient Flow [14]
(473) Minimax Estimation of Maximum Mean Discrepancy with Radial Kernels [465]
(474) A Kernel Two-Sample Test [154]
(475) Modular Block-diagonal Curvature Approximations for Feedforward Architectures [101]
(476) On the empirical estimation of integral probability metrics [435]
(477) On the importance of initialization and momentum in deep learning [443]
(478) Overcoming Classifier Imbalance for Long-tail Object Detection with Balanced Group Softmax [275]
(479) Pix2Vox++: Multi-scale Context-aware 3D Object Reconstruction from Single and Multiple Images [518]
(480) Radial basis functions for the multivariate interpolation of large scattered data sets [254]
(481) The Loss Surface of Deep and Wide Neural Networks [345]
(482) Training Neural Networks with Local Error Signals [348]

(483) Unbiased look at dataset bias [466]
(484) Understanding the Effective Receptive Field in Deep Convolutional Neural Networks [305]
(485) Visualizing the Loss Landscape of Neural Nets [266]
(486) Kernel Mean Embedding of Distributions: A Review and Beyonds [333]
(487) A Survey of Semantic Segmentation [458]
(488) A survey of the recent architectures of deep convolutional neural networks [222]
(489) A Survey on Metric Learning for Feature Vectors and Structured Data [30]
(490) A Survey on Multi-Task Learning [557]
(491) Automatic differentiation in machine learning: a survey [28]
(492) Optimization Methods for Large-Scale Machine Learning [51]
(493) Video Panoptic Segmentation [226]
(494) Understanding convolutional neural networks with a mathematical model [249]
(495) A Mathematical Theory of Deep Convolutional Neural Networks for Feature Extraction [507]
(496) A Modern Take on the Bias-Variance Tradeoff in Neural Networks [342]
(497) Benefits of Depth in Neural Networks [457]
(498) Diverse Neural Network Learns True Target Functions [516]
(499) Exact solutions to the nonlinear dynamics of learning in deep linear neural networks [402]
(500) Learning Neural Networks with Two Nonlinear Layers in Polynomial Time [147]
(501) Learning Polynomials with Neural Networks [10]
(502) Multilayer Feedforward Networks with a Non-Polynomial Activation Function Can Approximate Any Function [261]
(503) Neural Tangent Kernel: Convergence and Generalization in Neural Networks [205]
(504) On the Number of Linear Regions of Deep Neural Networks [329]
(505) Provable Bounds for Learning Some Deep Representations [17]
(506) Recovery Guarantees for One-hidden-layer Neural Networks [565]
(507) Shallow vs. Deep Sum-Product Networks [103]
(508) The Expressive Power of Neural Networks: A View from the Width [304]
(509) The Power of Depth for Feedforward Neural Networks [123]
(510) Optimization for Machine Learning [433]
(511) Characterizing the Universal Approximation Property [243]
(512) When and Why Are Deep Networks Better Than Shallow Ones? [318]
(513) Understanding Machine Learning: From Theory To Algorithms [414]
(514) Computer Vision: A Modern Approach [130]
(515) An Invitation to 3-D Vision: From Images to Geometric Models [308]
(516) Computer Vision - Algorithms and Applications [448]
(517) Introducing Monte Carlo methods with [389]
(518) Monte Carlo Sampling Methods Using Markov Chains and Their Applications [163]
(519) Applied Stochastic Processes [279]
(520) Deep Learning [150]
(521) Dive into Deep Learning [542]
(522) Machine learning - a probabilistic perspective [340]
(523) The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches [8]
(524) Adversarial Examples Are Not Bugs, They Are Features [201]
(525) Asymptotic Guarantees for Learning Generative Models with the Sliced-Wasserstein Distance [341]
(526) Attention is All you Need [474]
(527) Attention U-Net: Learning Where to Look for the Pancreas [354]

(528) Generalized Sliced Wasserstein Distances [236]
(529) Regularized Evolution for Image Classifier Architecture Search [375]
(530) Residual Networks Behave Like Ensembles of Relatively Shallow Networks [475]
(531) Swapout: Learning an ensemble of deep architectures [419]
(532) Mobilenetv2: Inverted residuals and linear bottlenecks [400]
(533) Searching for mobilenetv3 [186]
(534) Resnet in Resnet: Generalizing Residual Architectures [456]
(535) Residual Networks of Residual Networks: Multilevel Residual Networks [548]
(536) Generative adversarial nets [151]
(537) Bag of tricks for image classification with convolutional neural networks [172]
(538) Auto-Encoding Variational Bayes [231]
(539) Tutorial on Variational Autoencoders [108]
(540) InfoVAE: Information Maximizing Variational Autoencoders [561]
(541) Learning Representations by Maximizing Mutual Information in Variational Autoencoders [302]
(542) MAE: Mutual Posterior-Divergence Regularization for Variational AutoEncoders [307]
(543) Disentangling by Factorising [227]
(544) Elbo surgery: yet another way to carve up the variational evidence lower bound [185]
(545) Recent Advances in Autoencoder-Based Representation Learning [468]
(546) Variational Inference: A Review for Statisticians [44]
(547) Stochastic Backpropagation and Approximate Inference in Deep Generative Models [388]
(548) Learning Independent Features with Adversarial Nets for Non-linear ICA [54]
(549) Mutual Information Neural Estimation [29]
(550) Adversarially Learned Inference [116]
(551) Emergence of Invariance and Disentanglement in Deep Representations [1]
(552) Understanding the limitations of variational mutual information estimators [428]
(553) Information, divergence and risk for binary experiments [381]
(554) Extracting and composing robust features with denoising autoencoders [478]
(555) Representation Learning: A Review and New Perspectives [35]
(556) Deep High-Resolution Representation Learning for Visual Recognition [489]
(557) Stacked Generative Adversarial Networks [197]
(558) The IM Algorithm: A Variational Approach to Information Maximization [24]
(559) FineGAN: Unsupervised Hierarchical Disentanglement for Fine-Grained Object Generation and Discovery [418]
(560) Group equivariant convolutional networks [96]

REFERENCES

[1] A. Achille and Stefano Soatto. 2018. Emergence of Invariance and Disentanglement in Deep Representations. *2018 Information Theory and Applications Workshop (ITA)* (2018), 1–9.

[2] Eirikur Agustsson, Fabian Mentzer, Michael Tschannen, Lukas Cavigelli, Radu Timofte, Luca Benini, and Luc Van Gool. 2017. Soft-to-Hard Vector Quantization for End-to-End Learning Compressible Representations. arXiv:cs.LG/1704.00648

[3] Mohamad Faiz Ahmad, Nor Ashidi Mat Isa, Wei Hong Lim, and Koon Meng Ang. 2021. Differential evolution: A recent review based on state-of-the-art works. *Alexandria Engineering Journal* (2021).

[4] Jiwoon Ahn, Sunghyun Cho, and Suha Kwak. 2019. Weakly Supervised Learning of Instance Segmentation With Inter-Pixel Relations. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2019), 2204–2213.

[5] Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, and Mario Marchand. 2014. Domain-adversarial neural networks. *arXiv preprint arXiv:1412.4446* (2014).

[6] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. 2019. Optuna: A Next-generation Hyperparameter Optimization Framework. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (2019).

[7] Pablo Fernández Alcantarilla, Adrien Bartoli, and Andrew J. Davison. 2012. KAZE Features. In *ECCV*.

[8] M. Alom, T. Taha, Christopher Yakopcic, Stefan Westberg, M. Hasan, Brian C. Van Esesn, A. Awwal, and V. Asari. 2018. The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches. *ArXiv* abs/1803.01164 (2018).

[9] Jose M. Alvarez and Mathieu Salzmann. 2017. Compression-aware Training of Deep Networks. arXiv:cs.CV/1711.02638

[10] Alexandr Andoni, Rina Panigrahy, Gregory Valiant, and Li Zhang. 2014. Learning Polynomials with Neural Networks. In *ICML*.

[11] Jyoti Aneja, Alex Schwing, Jan Kautz, and Arash Vahdat. 2020. NCP-VAE: Variational autoencoders with noise contrastive priors. (2020).

[12] Sajid Anwar, Kyuyeon Hwang, and Wonyong Sung. 2017. Structured Pruning of Deep Convolutional Neural Networks. *ACM Journal on Emerging Technologies in Computing Systems (JETC)* 13 (2017), 1 – 18.

[13] Eric Arazo, Diego Ortego, Paul Albert, Noel E. O'Connor, and Kevin McGuinness. 2020. Pseudo-Labeling and Confirmation Bias in Deep Semi-Supervised Learning. *2020 International Joint Conference on Neural Networks (IJCNN)* (2020), 1–8.

[14] Michael Arbel, Anna Korba, Adil Salim, and Arthur Gretton. 2019. Maximum Mean Discrepancy Gradient Flow. *ArXiv* abs/1906.04370 (2019).

[15] Martin Arjovsky and Léon Bottou. 2017. Towards Principled Methods for Training Generative Adversarial Networks. arXiv:stat.ML/1701.04862

[16] Martin Arjovsky, Soumith Chintala, and Léon Bottou. 2017. Wasserstein GAN. arXiv:stat.ML/1701.07875

[17] Sanjeev Arora, Aditya Bhaskara, R. Ge, and Tengyu Ma. 2014. Provable Bounds for Learning Some Deep Representations. In *ICML*.

[18] Sanjeev Arora, Rong Ge, Yingyu Liang, Tengyu Ma, and Yi Zhang. 2017. Generalization and Equilibrium in Generative Adversarial Nets (GANs). arXiv:cs.LG/1703.00573

[19] Yuki M Asano, Christian Rupprecht, and Andrea Vedaldi. 2019. A critical analysis of self-supervision, or what we can learn from a single image. *arXiv preprint arXiv:1904.13132* (2019).

[20] Ben Athiwaratkun, Marc Finzi, Pavel Izmailov, and Andrew Gordon Wilson. 2018. There are many consistent explanations of unlabeled data: Why you should average. *arXiv preprint arXiv:1806.05594* (2018).

[21] Seyed Majid Azimi, Eleonora Vig, Reza Bahmanyar, Marco Körner, and Peter Reinartz. 2018. Towards multi-class object detection in unconstrained remote sensing imagery. In *Asian Conference on Computer Vision*. Springer, 150–165.

[22] Philip Bachman, R. Devon Hjelm, and William Buchwalter. 2019. Learning Representations by Maximizing Mutual Information Across Views. In *NeurIPS*.

[23] Bowen Baker, Otkrist Gupta, Nikhil Naik, and Ramesh Raskar. 2017. Designing Neural Network Architectures using Reinforcement Learning. *ArXiv* abs/1611.02167 (2017).

[24] D. Barber and F. Agakov. 2003. The IM Algorithm: A Variational Approach to Information Maximization. In *NIPS*.

[25] O. Barinova, V. Lempitsky, and Pushmeet Kohli. 2010. On Detection of Multiple Object Instances Using Hough Transforms. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34 (2010), 1773–1784.

[26] Shane Barratt and Rishi Sharma. 2018. A Note on the Inception Score. arXiv:stat.ML/1801.01973

[27] H. Bay, T. Tuytelaars, and L. Gool. 2006. SURF: Speeded Up Robust Features. In *ECCV*.

[28] Atilim Gunes Baydin, Barak A. Pearlmutter, Alexey Radul, and Jeffrey Mark Siskind. 2017. Automatic differentiation in machine learning: a survey. *ArXiv* abs/1502.05767 (2017).

[29] Mohamed Ishmael Belghazi, A. Baratin, Sai Rajeswar, Sherjil Ozair, Yoshua Bengio, R. Devon Hjelm, and Aaron C. Courville. 2018. Mutual Information Neural Estimation. In *ICML*.

[30] Aurélien Bellet, Amaury Habrard, and Marc Sebban. 2013. A Survey on Metric Learning for Feature Vectors and Structured Data. *ArXiv* abs/1306.6709 (2013).

[31] Irwan Bello. 2021. LambdaNetworks: Modeling Long-Range Interactions Without Attention. *ArXiv* abs/2102.08602 (2021).

[32] Irwan Bello, William Fedus, Xianzhi Du, Ekin Dogus Cubuk, A. Srinivas, Tsung-Yi Lin, Jonathon Shlens, and Barret Zoph. 2021. Revisiting ResNets: Improved Training and Scaling Strategies. *ArXiv* abs/2103.07579 (2021).
[33] Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and Jennifer Wortman Vaughan. 2010. A theory of learning from different domains. *Machine learning* 79, 1-2 (2010), 151–175.
[34] Shai Ben-David, John Blitzer, Koby Crammer, and Fernando Pereira. 2007. Analysis of representations for domain adaptation. In *Advances in neural information processing systems*. 137–144.
[35] Yoshua Bengio, Aaron C. Courville, and P. Vincent. 2013. Representation Learning: A Review and New Perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35 (2013), 1798–1828.
[36] Maxim Berman, A. Triki, and Matthew B. Blaschko. 2018. The Lovasz-Softmax Loss: A Tractable Surrogate for the Optimization of the Intersection-Over-Union Measure in Neural Networks. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2018), 4413–4421.
[37] David Berthelot, Nicholas Carlini, Ekin Dogus Cubuk, Alexey Kurakin, Kihyuk Sohn, Han Zhang, and Colin Raffel. 2020. ReMixMatch: Semi-Supervised Learning with Distribution Matching and Augmentation Anchoring. In *ICLR*.
[38] David Berthelot, Nicholas Carlini, Ian Goodfellow, Nicolas Papernot, Avital Oliver, and Colin A Raffel. 2019. Mixmatch: A holistic approach to semi-supervised learning. *Advances in Neural Information Processing Systems* 32 (2019).
[39] Simone Bianco, Remi Cadene, Luigi Celona, and Paolo Napoletano. 2018. Benchmark Analysis of Representative Deep Neural Network Architectures. *IEEE Access* 6 (2018), 64270–64277. https://doi.org/10.1109/access.2018.2877890
[40] Alberto Bietti and Julien Mairal. 2017. Invariance and stability of deep convolutional representations. *Advances in neural information processing systems* 30 (2017).
[41] Alberto Bietti and Julien Mairal. 2019. Group Invariance, Stability to Deformations, and Complexity of Deep Convolutional Representations. *J. Mach. Learn. Res.* 20 (2019), 25:1–25:49.
[42] Vighnesh Birodkar, Zhichao Lu, Siyang Li, Vivek Rathod, and Jonathan Huang. 2021. The surprising impact of mask-head architecture on novel class segmentation. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)* (2021), 6995–7005.
[43] Mikołaj Bińkowski, Dougal J. Sutherland, Michael Arbel, and Arthur Gretton. 2018. Demystifying MMD GANs. arXiv:stat.ML/1801.01401
[44] D. Blei, Alp Kucukelbir, and J. McAuliffe. 2016. Variational Inference: A Review for Statisticians. *J. Amer. Statist. Assoc.* 112 (2016), 859 – 877.
[45] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao. 2020. Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934* (2020).
[46] Marc Bocquet, Lin Wu, and Frédéric Chevallier. 2011. Bayesian design of control space for optimal assimilation of observations. Part I: Consistent multiscale formalism. *Quarterly Journal of the Royal Meteorological Society* 137, 658 (2011), 1340–1356.
[47] Navaneeth Bodla, Bharat Singh, Rama Chellappa, and Larry S. Davis. 2017. Soft-NMS — Improving Object Detection with One Line of Code. *2017 IEEE International Conference on Computer Vision (ICCV)* (2017), 5562–5570.
[48] Daniel Bolya, Chong Zhou, Fanyi Xiao, and Yong Jae Lee. 2019. YOLACT: Real-Time Instance Segmentation. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)* (2019), 9156–9165.
[49] Ali Borji, Ming-Ming Cheng, Huaizu Jiang, and Jun ru Li. 2019. Salient object detection: A survey. *Computational Visual Media* 5 (2019), 117–150.
[50] Léon Bottou. 2012. Stochastic Gradient Descent Tricks. In *Neural Networks: Tricks of the Trade*.
[51] Léon Bottou, Frank E. Curtis, and Jorge Nocedal. 2018. Optimization Methods for Large-Scale Machine Learning. *ArXiv* abs/1606.04838 (2018).
[52] Konstantinos Bousmalis, George Trigeorgis, Nathan Silberman, Dilip Krishnan, and Dumitru Erhan. 2016. Domain separation networks. In *Advances in neural information processing systems*. 343–351.
[53] Bert De Brabandere, Davy Neven, and Luc Van Gool. 2017. Semantic Instance Segmentation with a Discriminative Loss Function. *ArXiv* abs/1708.02551 (2017).
[54] Philemon Brakel and Yoshua Bengio. 2018. Learning Independent Features with Adversarial Nets for Non-linear ICA. *arXiv: Machine Learning* (2018).
[55] Hilton Bristow, Jack Valmadre, and Simon Lucey. 2015. Dense Semantic Correspondence where Every Pixel is a Classifier. arXiv:cs.CV/1505.04143
[56] Andrew Brock, Soham De, and Samuel L. Smith. 2021. Characterizing signal propagation to close the performance gap in unnormalized ResNets. *ArXiv* abs/2101.08692 (2021).
[57] Andrew Brock, Soham De, Samuel L. Smith, and Karen Simonyan. 2021. High-Performance Large-Scale Image Recognition Without Normalization. *ArXiv* abs/2102.06171 (2021).
[58] Andrew Brock, Jeff Donahue, and Karen Simonyan. 2018. Large scale GAN training for high fidelity natural image synthesis. *arXiv preprint arXiv:1809.11096* (2018).
[59] Sébastien Bubeck, Ronen Eldan, Yin Tat Lee, and Dan Mikulincer. 2020. Network size and weights size for memorization with two-layers neural networks. *ArXiv* abs/2006.02855 (2020).

- [60] Sébastien Bubeck, Yin Tat Lee, and Mohit Singh. 2015. A geometric alternative to Nesterov's accelerated gradient descent. *ArXiv* abs/1506.08187 (2015).
- [61] Mateusz Buda, Atsuto Maki, and Maciej A. Mazurowski. 2018. A systematic study of the class imbalance problem in convolutional neural networks. *Neural networks : the official journal of the International Neural Network Society* 106 (2018), 249–259.
- [62] Holger Caesar, Varun Bankiti, Alex H. Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. 2019. nuScenes: A multimodal dataset for autonomous driving. arXiv:cs.LG/1903.11027
- [63] Junxu Cao, Qi Chen, Jun Guo, and Ruichao Shi. 2020. Attention-guided Context Feature Pyramid Network for Object Detection. *ArXiv* abs/2005.11475 (2020).
- [64] Jiale Cao, Hisham Cholakkal, Rao Muhammad Anwer, Fahad Shahbaz Khan, Yanwei Pang, and Ling Shao. 2020. D2Det: Towards High Quality Object Detection and Instance Segmentation. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2020), 11482–11491.
- [65] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. 2020. End-to-End Object Detection with Transformers. *ArXiv* abs/2005.12872 (2020).
- [66] Mathilde Caron, Piotr Bojanowski, Armand Joulin, and Matthijs Douze. 2018. Deep Clustering for Unsupervised Learning of Visual Features. In *ECCV*.
- [67] Mathilde Caron, Hugo Touvron, Ishan Misra, Herv'e J'egou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. 2021. Emerging Properties in Self-Supervised Vision Transformers. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)* (2021), 9630–9640.
- [68] Ming-Fang Chang, John Lambert, Patsorn Sangkloy, Jagjeet Singh, Slawomir Bak, Andrew Hartnett, De Wang, Peter Carr, Simon Lucey, Deva Ramanan, and James Hays. 2019. Argoverse: 3D Tracking and Forecasting with Rich Maps. arXiv:cs.CV/1911.02620
- [69] Changan Chen, Frederick Tung, Naveen Vedula, and Greg Mori. 2018. Constraint-Aware Deep Neural Network Compression. In *ECCV*.
- [70] Guobin Chen, Wongun Choi, Xiang Yu, Tony X. Han, and Manmohan Chandraker. 2017. Learning Efficient Object Detection Models with Knowledge Distillation. In *NIPS*.
- [71] Kai Chen, Jiaqi Wang, Jiangmiao Pang, Yuhang Cao, Yu Xiong, Xiaoxiao Li, Shuyang Sun, Wansen Feng, Ziwei Liu, Jiarui Xu, Zheng Zhang, Dazhi Cheng, Chenchen Zhu, Tianheng Cheng, Qijie Zhao, Buyu Li, Xin Lu, Rui Zhu, Yue Wu, Jifeng Dai, Jingdong Wang, Jianping Shi, Wanli Ouyang, Chen Change Loy, and Dahua Lin. 2019. MMDetection: Open MMLab Detection Toolbox and Benchmark. *ArXiv* abs/1906.07155 (2019).
- [72] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. 2017. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587* (2017).
- [73] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. 2018. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. *ArXiv* abs/1802.02611 (2018).
- [74] Mickaël Chen, Thierry Artières, and Ludovic Denoyer. 2019. Unsupervised Object Segmentation by Redrawing. In *NeurIPS*.
- [75] Minmin Chen, Zhixiang Xu, Kilian Weinberger, and Fei Sha. 2012. Marginalized denoising autoencoders for domain adaptation. *arXiv preprint arXiv:1206.4683* (2012).
- [76] Qiang Chen, Yingming Wang, Tong Yang, X. Zhang, Jian Cheng, and Jian Sun. 2021. You Only Look One-level Feature. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2021), 13034–13043.
- [77] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey E. Hinton. 2020. A Simple Framework for Contrastive Learning of Visual Representations. *ArXiv* abs/2002.05709 (2020).
- [78] Ting Chen, Simon Kornblith, Kevin Swersky, Mohammad Norouzi, and Geoffrey E. Hinton. 2020. Big Self-Supervised Models are Strong Semi-Supervised Learners. *ArXiv* abs/2006.10029 (2020).
- [79] Ting Chen, Xiaohua Zhai, Marvin Ritter, Mario Lucic, and Neil Houlsby. 2018. Self-Supervised GANs via Auxiliary Rotation Loss. arXiv:cs.LG/1811.11212
- [80] Xi Chen, Yan Duan, Rein Houthooft, John Schulman, Ilya Sutskever, and Pieter Abbeel. 2016. InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets. arXiv:cs.LG/1606.03657
- [81] Xinlei Chen, Ross B. Girshick, Kaiming He, and Piotr Dollár. 2019. TensorMask: A Foundation for Dense Object Segmentation. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)* (2019), 2061–2069.
- [82] Xiangning Chen, Cihang Xie, Mingxing Tan, Li Zhang, Cho-Jui Hsieh, and Boqing Gong. 2021. Robust and Accurate Object Detection via Adversarial Learning. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2021), 16617–16626.
- [83] Yuhua Chen, Wen Li, Christos Sakaridis, Dengxin Dai, and Luc Van Gool. 2018. Domain adaptive faster r-cnn for object detection in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3339–3348.
- [84] Yukang Chen, Gaofeng Meng, Qian Zhang, Xinbang Zhang, Liangchen Song, Shiming Xiang, and Chunhong Pan. 2018. Joint Neural Architecture Search and Quantization. *ArXiv* abs/1811.09426 (2018).
- [85] Lu Chi, Zehuan Yuan, Yadong Mu, and Changhu Wang. 2020. Non-Local Neural Networks With Grouped Bilinear Attentional Transforms. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2020), 11801–11810.
- [86] Minsu Cho, Suha Kwak, Cordelia Schmid, and Jean Ponce. 2015. Unsupervised object discovery and localization in the wild: Part-based matching with bottom-up region proposals. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1201–1210.

[87] Junsuk Choe, Seungho Lee, and Hyunjung Shim. 2021. Attention-Based Dropout Layer for Weakly Supervised Single Object Localization and Semantic Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43 (2021), 4256–4271.

[88] François Chollet. 2016. Xception: Deep Learning with Depthwise Separable Convolutions. arXiv:cs.CV/1610.02357

[89] François Chollet. 2017. Xception: Deep Learning with Depthwise Separable Convolutions. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017), 1800–1807.

[90] Sumit Chopra, Suhrid Balakrishnan, and Raghuraman Gopalan. 2013. Dlid: Deep learning for domain adaptation by interpolating between domains. In *ICML workshop on challenges in representation learning*, Vol. 2.

[91] Christopher B. Choy, JunYoung Gwak, Silvio Savarese, and Manmohan Chandraker. 2016. Universal Correspondence Network. arXiv:cs.CV/1606.03558

[92] Patryk Chrabaszcz, Ilya Loshchilov, and Frank Hutter. 2017. A downsampled variant of imagenet as an alternative to the cifar datasets. *arXiv preprint arXiv:1707.08819* (2017).

[93] Djork-Arné Clevert, Thomas Unterthiner, and Sepp Hochreiter. 2015. Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs). arXiv:cs.LG/1511.07289

[94] Nadav Cohen, Or Sharir, and Amnon Shashua. 2015. On the Expressive Power of Deep Learning: A Tensor Analysis. *arXiv: Neural and Evolutionary Computing* (2015).

[95] Nadav Cohen and Amnon Shashua. 2016. Convolutional Rectifier Networks as Generalized Tensor Decompositions. *ArXiv* abs/1603.00162 (2016).

[96] Taco S. Cohen and Max Welling. 2016. Group Equivariant Convolutional Networks. arXiv:cs.LG/1602.07576

[97] Jean-Baptiste Cordonnier, Andreas Loukas, and Martin Jaggi. 2020. On the Relationship between Self-Attention and Convolutional Layers. *ArXiv* abs/1911.03584 (2020).

[98] Ekin D Cubuk, Barret Zoph, Dandelion Mane, Vijay Vasudevan, and Quoc V Le. 2018. Autoaugment: Learning augmentation policies from data. *arXiv preprint arXiv:1805.09501* (2018).

[99] Ekin D Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V Le. 2020. Randaugment: Practical automated data augmentation with a reduced search space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. 702–703.

[100] Yin Cui, Yang Song, Chen Sun, Andrew Howard, and Serge Belongie. 2018. Large scale fine-grained categorization and domain-specific transfer learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4109–4118.

[101] Felix Dangel, Stefan Harmeling, and Philipp Hennig. 2020. Modular Block-diagonal Curvature Approximations for Feedforward Architectures. In *AISTATS*.

[102] Swagatam Das and Ponnuthurai Nagaratnam Suganthan. 2011. Differential Evolution: A Survey of the State-of-the-Art. *IEEE Transactions on Evolutionary Computation* 15 (2011), 4–31.

[103] Olivier Delalleau and Yoshua Bengio. 2011. Shallow vs. Deep Sum-Product Networks. In *NIPS*.

[104] Manuel Fernández Delgado, Eva Cernadas, Senén Barro, and Dinani Gomes Amorim. 2014. Do we need hundreds of classifiers to solve real world classification problems? *J. Mach. Learn. Res.* 15 (2014), 3133–3181.

[105] Qi Deng, Yi Cheng, and Guanghui Lan. 2018. Optimal Adaptive and Accelerated Stochastic Gradient Descent. *ArXiv* abs/1810.00553 (2018).

[106] Emily Denton, Soumith Chintala, Arthur Szlam, and Rob Fergus. 2015. Deep Generative Image Models using a Laplacian Pyramid of Adversarial Networks. arXiv:cs.CV/1506.05751

[107] Terrance DeVries and Graham W Taylor. 2017. Improved regularization of convolutional neural networks with cutout. *arXiv preprint arXiv:1708.04552* (2017).

[108] C. Doersch. 2016. Tutorial on Variational Autoencoders. *ArXiv* abs/1606.05908 (2016).

[109] Carl Doersch, Abhinav Kumar Gupta, and Alexei A. Efros. 2015. Unsupervised Visual Representation Learning by Context Prediction. *2015 IEEE International Conference on Computer Vision (ICCV)* (2015), 1422–1430.

[110] Piotr Dollár, Mannat Singh, and Ross B. Girshick. 2021. Fast and Accurate Model Scaling. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2021), 924–932.

[111] Jeff Donahue, Philipp Krähenbühl, and Trevor Darrell. 2016. Adversarial Feature Learning. arXiv:cs.LG/1605.09782

[112] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. 2021. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *ArXiv* abs/2010.11929 (2021).

[113] Timothy Dozat. 2016. Incorporating Nesterov Momentum into Adam.

[114] S. Du, Xiyu Zhai, B. Póczos, and A. Singh. 2019. Gradient Descent Provably Optimizes Over-parameterized Neural Networks. *ArXiv* abs/1810.02054 (2019).

[115] Xianzhi Du, Tsung-Yi Lin, Pengchong Jin, Golnaz Ghiasi, Mingxing Tan, Yin Cui, Quoc V. Le, and Xiaodan Song. 2020. SpineNet: Learning Scale-Permuted Backbone for Recognition and Localization. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2020), 11589–11598.

[116] Vincent Dumoulin, Ishmael Belghazi, Ben Poole, Alex Lamb, Martín Arjovsky, Olivier Mastropietro, and Aaron C. Courville. 2017. Adversarially Learned Inference. *ArXiv* abs/1606.00704 (2017).

[117] Vincent Dumoulin and Francesco Visin. 2016. A guide to convolution arithmetic for deep learning. *ArXiv* abs/1603.07285 (2016).

[118] Ionut Cosmin Duta, Li Liu, Fan Zhu, and Ling Shao. 2020. Pyramidal Convolution: Rethinking Convolutional Neural Networks for Visual Recognition. arXiv:cs.CV/2006.11538

[119] Nikita Dvornik, Julien Mairal, and Cordelia Schmid. 2018. Modeling Visual Context is Key to Augmenting Object Detection Datasets. In *ECCV*.

[120] Gintare Karolina Dziugaite, Daniel M. Roy, and Zoubin Ghahramani. 2015. Training generative neural networks via Maximum Mean Discrepancy optimization. arXiv:stat.ML/1505.03906

[121] Patrick Ebel, Anastasiia Mishchuk, Kwang Moo Yi, Pascal Fua, and Eduard Trulls. 2019. Beyond Cartesian Representations for Local Descriptors. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)* (2019), 253–262.

[122] David Eigen, Jason Rolfe, Rob Fergus, and Yann LeCun. 2013. Understanding Deep Architectures using a Recursive Convolutional Network. arXiv:cs.LG/1312.1847

[123] Ronen Eldan and Ohad Shamir. 2016. The Power of Depth for Feedforward Neural Networks. In *COLT*.

[124] Russell A. Epstein and Chris I. Baker. 2019. Scene Perception in the Human Brain. *Annual review of vision science* (2019).

[125] Junsong Fan, Zhaoxiang Zhang, and Tieniu Tan. 2020. CIAN: Cross-Image Affinity Net for Weakly Supervised Semantic Segmentation. In *AAAI*.

[126] Haoshu Fang, Jianhua Sun, Runzhong Wang, Minghao Gou, Yong-Lu Li, and Cewu Lu. 2019. InstaBoost: Boosting Instance Segmentation via Probability Map Guided Copy-Pasting. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)* (2019), 682–691.

[127] Michael Figurnov, Aijan Ibraimova, Dmitry Vetrov, and Pushmeet Kohli. 2015. PerforatedCNNs: Acceleration through Elimination of Redundant Convolutions. arXiv:cs.CV/1504.08362

[128] Graham D. Finlayson. 2018. Colour and illumination in computer vision. *Interface Focus* 8 (2018).

[129] Pierre Foret, Ariel Kleiner, Hossein Mobahi, and Behnam Neyshabur. 2021. Sharpness-Aware Minimization for Efficiently Improving Generalization. *ArXiv* abs/2010.01412 (2021).

[130] D. Forsyth and J. Ponce. 2002. Computer Vision: A Modern Approach.

[131] Yaroslav Ganin and Victor Lempitsky. 2015. Unsupervised domain adaptation by backpropagation. In *International conference on machine learning*. 1180–1189.

[132] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. 2016. Domain-adversarial training of neural networks. *The Journal of Machine Learning Research* 17, 1 (2016), 2096–2030.

[133] Yang Gao, Oscar Beijbom, Ning Zhang, and Trevor Darrell. 2016. Compact Bilinear Pooling. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016), 317–326.

[134] Patrick Garrigan and Philip J. Kellman. 2008. Perceptual learning depends on perceptual constancy. *Proceedings of the National Academy of Sciences* 105 (2008), 2248 – 2253.

[135] Xavier Gastaldi. 2017. Shake-shake regularization. *arXiv preprint arXiv:1705.07485* (2017).

[136] Zheng Ge, Songtao Liu, Feng Wang, Zeming Li, and Jian Sun. 2021. YOLOX: Exceeding YOLO Series in 2021. *ArXiv* abs/2107.08430 (2021).

[137] Robert Geirhos, Carlos R. Medina Temme, Jonas Rauber, Heiko Herbert Schütt, Matthias Bethge, and Felix Wichmann. 2018. Generalisation in humans and deep neural networks. *ArXiv* abs/1808.08750 (2018).

[138] Golnaz Ghiasi, Yin Cui, A. Srinivas, Rui Qian, Tsung-Yi Lin, Ekin Dogus Cubuk, Quoc V. Le, and Barret Zoph. 2021. Simple Copy-Paste is a Strong Data Augmentation Method for Instance Segmentation. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2021), 2917–2927.

[139] Golnaz Ghiasi, Tsung-Yi Lin, and Quoc V Le. 2018. Dropblock: A regularization method for convolutional networks. In *Advances in Neural Information Processing Systems*. 10727–10737.

[140] Golnaz Ghiasi, Tsung-Yi Lin, Ruoming Pang, and Quoc V. Le. 2019. NAS-FPN: Learning Scalable Feature Pyramid Architecture for Object Detection. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2019), 7029–7038.

[141] Spyros Gidaris, Praveer Singh, and Nikos Komodakis. 2018. Unsupervised Representation Learning by Predicting Image Rotations. *ArXiv* abs/1803.07728 (2018).

[142] Ross Girshick. 2015. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*. 1440–1448.

[143] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 580–587.

[144] Ross B. Girshick. 2015. Fast R-CNN. *2015 IEEE International Conference on Computer Vision (ICCV)* (2015), 1440–1448.

[145] Xavier Glorot and Yoshua Bengio. 2010. Understanding the difficulty of training deep feedforward neural networks. In *AISTATS*.

[146] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. 2011. Deep Sparse Rectifier Neural Networks. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics (Proceedings of Machine Learning Research)*, Geoffrey Gordon, David Dunson, and Miroslav Dudík (Eds.), Vol. 15. PMLR, Fort Lauderdale, FL, USA, 315–323. http://proceedings.mlr.press/v15/glorot11a.html

[147] Surbhi Goel and Adam R. Klivans. 2019. Learning Neural Networks with Two Nonlinear Layers in Polynomial Time. In *COLT*.

[148] Mingming Gong, Yanwu Xu, Chunyuan Li, Kun Zhang, and Kayhan Batmanghelich. 2019. Twin Auxiliary Classifiers GAN. arXiv:cs.LG/1907.02690

[149] Xinyu Gong, Shiyu Chang, Yifan Jiang, and Zhangyang Wang. 2019. AutoGAN: Neural Architecture Search for Generative Adversarial Networks. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)* (2019), 3223–3233.

[150] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning*. MIT Press. http://www.deeplearningbook.org.

[151] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. In *Advances in neural information processing systems*. 2672–2680.

[152] Ian J Goodfellow, Yaroslav Bulatov, Julian Ibarz, Sacha Arnoud, and Vinay Shet. 2013. Multi-digit number recognition from street view imagery using deep convolutional neural networks. *arXiv preprint arXiv:1312.6082* (2013).

[153] Ian J. Goodfellow, David Warde-Farley, Mehdi Mirza, Aaron Courville, and Yoshua Bengio. 2013. Maxout Networks. arXiv:stat.ML/1302.4389

[154] Arthur Gretton, Karsten M. Borgwardt, Malte J. Rasch, Bernhard Schölkopf, and Alexander J. Smola. 2012. A Kernel Two-Sample Test. *J. Mach. Learn. Res.* 13 (2012), 723–773.

[155] Roger B. Grosse and J. Martens. 2016. A Kronecker-factored approximate Fisher matrix for convolution layers. In *ICML*.

[156] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron Courville. 2017. Improved Training of Wasserstein GANs. arXiv:cs.LG/1704.00028

[157] Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q. Weinberger. 2017. On Calibration of Modern Neural Networks. *ArXiv* abs/1706.04599 (2017).

[158] Hongyu Guo, Yongyi Mao, and Richong Zhang. 2019. MixUp as Locally Linear Out-Of-Manifold Regularization. In *AAAI*.

[159] Dongyoon Han, Jiwhan Kim, and Junmo Kim. 2016. Deep Pyramidal Residual Networks. arXiv:cs.CV/1610.02915

[160] Dongyoon Han, Sangdoo Yun, Byeongho Heo, and Young Joon Yoo. 2021. Rethinking Channel Dimensions for Efficient Model Design. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2021), 732–741.

[161] Kai Han, Yunhe Wang, Qi Tian, Jianyuan Guo, Chunjing Xu, and Chang Xu. 2019. GhostNet: More Features from Cheap Operations. arXiv:cs.CV/1911.11907

[162] Kai Han, Yunhe Wang, Qi Tian, Jianyuan Guo, Chunjing Xu, and Chang Xu. 2020. GhostNet: More Features From Cheap Operations. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2020), 1577–1586.

[163] W. D. Hastings. 1970. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika* 57 (1970), 97–109.

[164] J. Haxby, M. I. Gobbini, M. Furey, A. Ishai, J. L. Schouten, and P. Pietrini. 2001. Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex. *Science* 293 (2001), 2425 – 2430.

[165] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross B. Girshick. 2020. Momentum Contrast for Unsupervised Visual Representation Learning. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2020), 9726–9735.

[166] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. 2017. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*. 2961–2969.

[167] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2014. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *Lecture Notes in Computer Science* (2014), 346–361. https://doi.org/10.1007/978-3-319-10578-9_23

[168] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. Deep Residual Learning for Image Recognition. arXiv:cs.CV/1512.03385

[169] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. *2015 IEEE International Conference on Computer Vision (ICCV)* (2015), 1026–1034.

[170] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Identity Mappings in Deep Residual Networks. arXiv:cs.CV/1603.05027

[171] Tong He, Zhi Zhang, Hang Zhang, Zhongyue Zhang, Junyuan Xie, and Mu Li. 2019. Bag of Tricks for Image Classification with Convolutional Neural Networks. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2019), 558–567.

[172] Tong He, Zhi Zhang, Hang Zhang, Zhongyue Zhang, Junyuan Xie, and Mu Li. 2019. Bag of tricks for image classification with convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 558–567.

[173] Yihui He, Xiangyu Zhang, and Jian Sun. 2017. Channel Pruning for Accelerating Very Deep Neural Networks. *2017 IEEE International Conference on Computer Vision (ICCV)* (2017), 1398–1406.

[174] Zhenwei He and Lei Zhang. 2019. Multi-adversarial faster-rcnn for unrestricted object detection. In *Proceedings of the IEEE International Conference on Computer Vision*. 6668–6677.

[175] Olivier J. Hénaff, A. Srinivas, Jeffrey De Fauw, Ali Razavi, Carl Doersch, S. M. Ali Eslami, and Aäron van den Oord. 2020. Data-Efficient Image Recognition with Contrastive Predictive Coding. *ArXiv* abs/1905.09272 (2020).

[176] Dan Hendrycks and Thomas G. Dietterich. 2019. Benchmarking Neural Network Robustness to Common Corruptions and Perturbations. *ArXiv* abs/1903.12261 (2019).

[177] Dan Hendrycks and Kevin Gimpel. 2017. A Baseline for Detecting Misclassified and Out-of-Distribution Examples in Neural Networks. *ArXiv* abs/1610.02136 (2017).

[178] Dan Hendrycks, Xiaoyuan Liu, Eric Wallace, Adam Dziedzic, Rishabh Krishnan, and Dawn Song. 2020. Pretrained transformers improve out-of-distribution robustness. *arXiv preprint arXiv:2004.06100* (2020).

[179] Dan Hendrycks, Xiaoyuan Liu, Eric Wallace, Adam Dziedzic, Rishabh Krishnan, and Dawn Xiaodong Song. 2020. Pretrained Transformers Improve Out-of-Distribution Robustness. In *ACL*.

[180] Dan Hendrycks, Mantas Mazeika, and Thomas G. Dietterich. 2019. Deep Anomaly Detection with Outlier Exposure. *ArXiv* abs/1812.04606 (2019).

[181] Dan Hendrycks, Norman Mu, Ekin D Cubuk, Barret Zoph, Justin Gilmer, and Balaji Lakshminarayanan. 2019. Augmix: A simple data processing method to improve robustness and uncertainty. *arXiv preprint arXiv:1912.02781* (2019).

[182] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. 2017. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. arXiv:cs.LG/1706.08500

[183] Daniel Ho, Eric Liang, Xi Chen, Ion Stoica, and Pieter Abbeel. 2019. Population based augmentation: Efficient learning of augmentation policy schedules. In *International Conference on Machine Learning*. 2731–2741.

[184] Elad Hoffer, Itay Hubara, and Daniel Soudry. 2017. Train longer, generalize better: closing the generalization gap in large batch training of neural networks. *ArXiv* abs/1705.08741 (2017).

[185] Matthew D Hoffman and Matthew J Johnson. [n.d.]. Elbo surgery: yet another way to carve up the variational evidence lower bound.

[186] Andrew Howard, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan, Weijun Wang, Yukun Zhu, Ruoming Pang, Vijay Vasudevan, et al. 2019. Searching for mobilenetv3. In *Proceedings of the IEEE International Conference on Computer Vision*. 1314–1324.

[187] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. 2017. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv:cs.CV/1704.04861

[188] Jie Hu, Li Shen, Samuel Albanie, Gang Sun, and Andrea Vedaldi. 2018. Gather-Excite: Exploiting Feature Context in Convolutional Neural Networks. In *NeurIPS*.

[189] Jie Hu, Li Shen, Samuel Albanie, Gang Sun, and Enhua Wu. 2017. Squeeze-and-Excitation Networks. arXiv:cs.CV/1709.01507

[190] Jianming Hu, Xiyang Zhi, Tianjun Shi, Wei Zhang, Yang Cui, and Shenggang Zhao. 2021. PAG-YOLO: A Portable Attention-Guided YOLO Network for Small Ship Detection. *Remote. Sens.* 13 (2021), 3059.

[191] Ronghang Hu, Piotr Dollár, Kaiming He, Trevor Darrell, and Ross B. Girshick. 2018. Learning to Segment Every Thing. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2018), 4233–4241.

[192] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. 2016. Densely Connected Convolutional Networks. arXiv:cs.CV/1608.06993

[193] Gao Huang, Zhuang Liu, and Kilian Q. Weinberger. 2017. Densely Connected Convolutional Networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017), 2261–2269.

[194] Gao Huang, Yu Sun, Zhuang Liu, Daniel Sedra, and Kilian Q. Weinberger. 2016. Deep Networks with Stochastic Depth. In *ECCV*.

[195] Gao Huang, Yu Sun, Zhuang Liu, Daniel Sedra, and Kilian Q Weinberger. 2016. Deep networks with stochastic depth. In *European conference on computer vision*. Springer, 646–661.

[196] Jonathan Huang, Vivek Rathod, Chen Sun, Menglong Zhu, Anoop Korattikara, Alireza Fathi, Ian Fischer, Zbigniew Wojna, Yang Song, Sergio Guadarrama, et al. 2017. Speed/accuracy trade-offs for modern convolutional object detectors. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 7310–7311.

[197] X. Huang, Y. Li, Omid Poursaeed, J. Hopcroft, and Serge J. Belongie. 2017. Stacked Generative Adversarial Networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017), 1866–1875.

[198] Xin Huang, Xinxin Wang, Wenyu Lv, Xiaying Bai, Xiang Long, Kaipeng Deng, Qingqing Dang, Shumin Han, Qiwen Liu, Xiaoguang Hu, Dianhai Yu, Yanjun Ma, and Osamu Yoshie. 2021. PP-YOLOv2: A Practical Object Detector. *ArXiv* abs/2104.10419 (2021).

[199] Zilong Huang, Xinggang Wang, Lichao Huang, Chang Huang, Yunchao Wei, Humphrey Shi, and Wenyu Liu. 2019. CCNet: Criss-Cross Attention for Semantic Segmentation. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)* (2019), 603–612.

[200] Zeyi Huang, Yang Zou, B. V. K. Vijaya Kumar, and Dong Huang. 2020. Comprehensive Attention Self-Distillation for Weakly-Supervised Object Detection. *ArXiv* abs/2010.12023 (2020).

[201] Andrew Ilyas, Shibani Santurkar, D. Tsipras, L. Engstrom, B. Tran, and A. Madry. 2019. Adversarial Examples Are Not Bugs, They Are Features. In *NeurIPS*.

[202] Sergey Ioffe and Christian Szegedy. 2015. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *ArXiv* abs/1502.03167 (2015).

[203] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. 2016. Image-to-Image Translation with Conditional Adversarial Networks. arXiv:cs.CV/1611.07004

[204] Pavel Izmailov, Dmitrii Podoprikhin, T. Garipov, Dmitry P. Vetrov, and Andrew Gordon Wilson. 2018. Averaging Weights Leads to Wider Optima and Better Generalization. In *UAI*.

[205] Arthur Jacot, Franck Gabriel, and Clément Hongler. 2018. Neural Tangent Kernel: Convergence and Generalization in Neural Networks. In *NeurIPS*.

[206] Max Jaderberg, Karen Simonyan, Andrew Zisserman, et al. 2015. Spatial transformer networks. In *Advances in neural information processing systems*. 2017–2025.

[207] Max Jaderberg, Karen Simonyan, Andrew Zisserman, and Koray Kavukcuoglu. 2015. Spatial Transformer Networks. In *NIPS*.

[208] Stanisław Jastrzębski, Zachary Kenton, Devansh Arpit, Nicolas Ballas, Asja Fischer, Yoshua Bengio, and Amos Storkey. 2017. Three factors influencing minima in sgd. *arXiv preprint arXiv:1711.04623* (2017).

[209] Stanislaw Jastrzebski, Zachary Kenton, Devansh Arpit, Nicolas Ballas, Asja Fischer, Yoshua Bengio, and Amos J. Storkey. 2017. Three Factors Influencing Minima in SGD. *ArXiv* abs/1711.04623 (2017).

[210] Sangryul Jeon, Seungryong Kim, Dongbo Min, and Kwanghoon Sohn. 2018. PARN: Pyramidal Affine Regression Networks for Dense Semantic Correspondence. arXiv:cs.CV/1807.02939

[211] Xu Ji, Andrea Vedaldi, and João F. Henriques. 2019. Invariant Information Clustering for Unsupervised Image Classification and Segmentation. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)* (2019), 9864–9873.

[212] Kang jik Kim and Hee Seok Lee. 2020. Probabilistic Anchor Assignment with IoU Prediction for Object Detection. *ArXiv* abs/2007.08103 (2020).

[213] X. Jin, Jiang Wang, J. Slocum, Ming-Hsuan Yang, Shengyang Dai, S. Yan, and Jiashi Feng. 2019. RC-DARTS: Resource Constrained Differentiable Architecture Search. *ArXiv* abs/1912.12814 (2019).

[214] Eric R Kandel, James H Schwartz, Thomas M Jessell, Department of Biochemistry, Molecular Biophysics Thomas Jessell, Steven Siegelbaum, and AJ Hudspeth. 2000. *Principles of neural science*. Vol. 4. McGraw-hill New York.

[215] Guoliang Kang, Xuanyi Dong, Liang Zheng, and Yi Yang. 2017. PatchShuffle Regularization. *ArXiv* abs/1707.07103 (2017).

[216] Harini Kannan, Alexey Kurakin, and Ian J. Goodfellow. 2018. Adversarial Logit Pairing. *ArXiv* abs/1803.06373 (2018).

[217] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. 2017. Progressive Growing of GANs for Improved Quality, Stability, and Variation. arXiv:cs.NE/1710.10196

[218] Elias Kassapis, Georgi Dikov, Deepak K Gupta, and Cedric Nugteren. 2020. Calibrated Adversarial Refinement for Multimodal Semantic Segmentation. *arXiv preprint arXiv:2006.13144* (2020).

[219] Kenji Kawaguchi, Leslie Pack Kaelbling, and Yoshua Bengio. 2017. Generalization in deep learning. *arXiv preprint arXiv:1710.05468* (2017).

[220] Alex Kendall, Vijay Badrinarayanan, and Roberto Cipolla. 2015. Bayesian segnet: Model uncertainty in deep convolutional encoder-decoder architectures for scene understanding. *arXiv preprint arXiv:1511.02680* (2015).

[221] Hoel Kervadec, José Dolz, Shanshan Wang, Éric Granger, and Ismail Ben Ayed. 2020. Bounding boxes for weakly supervised segmentation: Global constraints get close to full supervision. In *MIDL*.

[222] Asifullah Khan, Anabia Sohail, Umme Zahoora, and Aqsa Saeed Qureshi. 2020. A survey of the recent architectures of deep convolutional neural networks. *Artificial Intelligence Review* (2020), 1 – 62.

[223] Salman Hameed Khan, Muzammal Naseer, Munawar Hayat, Syed Waqas Zamir, Fahad Shahbaz Khan, and Mubarak Shah. 2022. Transformers in Vision: A Survey. *ACM Computing Surveys (CSUR)* (2022).

[224] Anna Khoreva, Rodrigo Benenson, Jan Hendrik Hosang, Matthias Hein, and Bernt Schiele. 2017. Simple Does It: Weakly Supervised Instance and Semantic Segmentation. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017), 1665–1674.

[225] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. 2020. Supervised Contrastive Learning. *ArXiv* abs/2004.11362 (2020).

[226] Dahun Kim, Sanghyun Woo, Joonyoung Lee, and In So Kweon. 2020. Video Panoptic Segmentation. *ArXiv* abs/2006.11339 (2020).

[227] Hyunjik Kim and A. Mnih. 2018. Disentangling by Factorising. In *ICML*.

[228] Seungryong Kim, Dongbo Min, Bumsub Ham, Sangryul Jeon, Stephen Lin, and Kwanghoon Sohn. 2017. FCSS: Fully Convolutional Self-Similarity for Dense Semantic Correspondence. arXiv:cs.CV/1702.00926

[229] Diederik P. Kingma and Jimmy Ba. 2014. Adam: A Method for Stochastic Optimization. arXiv:cs.LG/1412.6980

[230] Diederik P Kingma and Max Welling. 2013. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114* (2013).

[231] Diederik P. Kingma and M. Welling. 2014. Auto-Encoding Variational Bayes. *CoRR* abs/1312.6114 (2014).

[232] Alexander Kirillov, Kaiming He, Ross Girshick, Carsten Rother, and Piotr Dollár. 2019. Panoptic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 9404–9413.

[233] Mate Kisantal, Zbigniew Wojna, Jakub Murawski, Jacek Naruniec, and Kyunghyun Cho. 2019. Augmentation for small object detection. *ArXiv* abs/1902.07296 (2019).

[234] Günter Klambauer, Thomas Unterthiner, Andreas Mayr, and Sepp Hochreiter. 2017. Self-Normalizing Neural Networks. *ArXiv* abs/1706.02515 (2017).

[235] Naveen Kodali, James Hays, Jacob D. Abernethy, and Zsolt Kira. 2018. On Convergence and Stability of GANs. *arXiv: Artificial Intelligence* (2018).

[236] Soheil Kolouri, Kimia Nadjahi, Umut Simsekli, R. Badeau, and G. Rohde. 2019. Generalized Sliced Wasserstein Distances. In *NeurIPS*.

[237] Tao Kong, Fuchun Sun, Wen bing Huang, and Huaping Liu. 2018. Deep Feature Pyramid Reconfiguration for Object Detection. In *ECCV*.

[238] Tao Kong, Fuchun Sun, Huaping Liu, Yuning Jiang, Lei Li, and Jianbo Shi. 2020. FoveaBox: Beyound Anchor-Based Object Detection. *IEEE Transactions on Image Processing* 29 (2020), 7389–7398.

[239] Simon Kornblith, Jonathon Shlens, and Quoc V. Le. 2019. Do Better ImageNet Models Transfer Better? *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2019), 2656–2666.

[240] Simon Kornblith, Jonathon Shlens, and Quoc V Le. 2019. Do better imagenet models transfer better?. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2661–2671.

[241] Z. Kourtzi, A. Tolias, C. Altmann, M. Augath, and N. Logothetis. 2003. Integration of Local Features into Global Shapes Monkey and Human fMRI Studies. *Neuron* 37 (2003), 333–346.

[242] Philipp Krähenbühl and Vladlen Koltun. 2011. Efficient inference in fully connected crfs with gaussian edge potentials. *Advances in neural information processing systems* 24 (2011).

[243] Anastasis Kratsios. 2019. Characterizing the Universal Approximation Property.

[244] Jonathan Krause, Michael Stark, Jia Deng, and Li Fei-Fei. 2013. 3d object representations for fine-grained categorization. In *Proceedings of the IEEE international conference on computer vision workshops*. 554–561.

[245] N. Kriegeskorte, Marieke Mur, D. Ruff, R. Kiani, J. Bodurka, H. Esteky, K. Tanaka, and P. Bandettini. 2008. Matching Categorical Object Representations in Inferior Temporal Cortex of Man and Monkey. *Neuron* 60 (2008), 1126–1141.

[246] Alex Krizhevsky, Geoffrey Hinton, et al. 2009. *Learning multiple layers of features from tiny images*. Technical Report. Citeseer.

[247] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. [n.d.]. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*. 2012.

[248] Ananya Kumar, Percy Liang, and Tengyu Ma. 2019. Verified Uncertainty Calibration. *ArXiv* abs/1909.10155 (2019).

[249] C.-C. Jay Kuo. 2016. Understanding convolutional neural networks with a mathematical model. *ArXiv* abs/1609.04112 (2016).

[250] Axel Barroso Laguna, Edgar Riba, Daniel Ponsa, and Krystian Mikolajczyk. 2019. Key.Net: Keypoint Detection by Handcrafted and Learned CNN Filters. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)* (2019), 5835–5843.

[251] Samuli Laine and Timo Aila. 2016. Temporal ensembling for semi-supervised learning. *arXiv preprint arXiv:1610.02242* (2016).

[252] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. 2017. Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles. In *NIPS*.

[253] Hei Law and Jia Deng. 2018. CornerNet: Detecting Objects as Paired Keypoints. *ArXiv* abs/1808.01244 (2018).

[254] D. Lazzaro and L. Montefusco. 2002. Radial basis functions for the multivariate interpolation of large scattered data sets.

[255] Dong-Hyun Lee. 2013. Pseudo-Label : The Simple and Efficient Semi-Supervised Learning Method for Deep Neural Networks.

[256] Youngwan Lee and Jongyoul Park. 2020. CenterMask: Real-Time Anchor-Free Instance Segmentation. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2020), 13903–13912.

[257] Youngwan Lee, Joong won Hwang, Sangrok Lee, Yuseok Bae, and Jongyoul Park. 2019. An Energy and GPU-Computation Efficient Backbone Network for Real-Time Object Detection. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (2019), 752–760.

[258] Joffrey L. Leevy, Taghi M. Khoshgoftaar, Richard A. Bauder, and Naeem Seliya. 2018. A survey on addressing high-class imbalance in big data. *Journal of Big Data* 5 (2018), 1–30.

[259] Karel Lenc and Andrea Vedaldi. 2015. Understanding image representations by measuring their equivariance and equivalence. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2015), 991–999.

[260] Karel Lenc and Andrea Vedaldi. 2015. Understanding image representations by measuring their equivariance and equivalence. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 991–999.

[261] M. Leshno, Vladimir Ya. Lin, A. Pinkus, and S. Schocken. 1993. Multilayer Feedforward Networks with a Non-Polynomial Activation Function Can Approximate Any Function. In *Neural Networks*.

[262] Aitor Lewkowycz, Yasaman Bahri, Ethan Dyer, Jascha Sohl-Dickstein, and Guy Gur-Ari. 2020. The large learning rate phase of deep learning: the catapult mechanism. *arXiv preprint arXiv:2003.02218* (2020).

[263] Ang Li, Ola Spyra, Sagi Perel, Valentin Dalibard, Max Jaderberg, Chenjie Gu, David Budden, Tim Harley, and Pramod Gupta. 2019. A generalized framework for population based training. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 1791–1799.

[264] Chongxuan Li, Kun Xu, Jun Zhu, and Bo Zhang. 2017. Triple Generative Adversarial Nets. arXiv:cs.LG/1703.02291

[265] Chun-Liang Li, Wei-Cheng Chang, Yu Cheng, Yiming Yang, and Barnabás Póczos. 2017. MMD GAN: Towards Deeper Understanding of Moment Matching Network. arXiv:cs.LG/1705.08584

[266] Hao Li, Zheng Xu, Gavin Taylor, and Tom Goldstein. 2018. Visualizing the Loss Landscape of Neural Nets. In *NeurIPS*.

[267] Liam Li, Kevin Jamieson, Afshin Rostamizadeh, Ekaterina Gonina, Jonathan Ben-Tzur, Moritz Hardt, Benjamin Recht, and Ameet Talwalkar. 2020. A system for massively parallel hyperparameter tuning. *Proceedings of Machine Learning and Systems* 2 (2020), 230–246.

[268] Qizhu Li, Anurag Arnab, and Philip HS Torr. 2018. Weakly-and semi-supervised panoptic segmentation. In *Proceedings of the European Conference on Computer Vision (ECCV)*. 102–118.

[269] Xiang Li, Wenhai Wang, Xiaolin Hu, and Jian Yang. 2019. Selective Kernel Networks. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2019), 510–519.
[270] Yanwei Li, Xinze Chen, Zheng Zhu, Lingxi Xie, Guan Huang, Dalong Du, and Xingang Wang. 2019. Attention-guided unified network for panoptic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 7026–7035.
[271] Yanghao Li, Yuntao Chen, Naiyan Wang, and Zhaoxiang Zhang. 2019. Scale-Aware Trident Networks for Object Detection. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)* (2019), 6053–6062.
[272] Yiting Li, Haisong Huang, Qingsheng Xie, Liguo Yao, and Qipeng Chen. 2018. Research on a Surface Defect Detection Algorithm Based on MobileNet-SSD. *Applied Sciences* (2018).
[273] Yi Li, Zhanghui Kuang, Liyang Liu, Yimin Chen, and Wayne Zhang. 2021. Pseudo-mask Matters in Weakly-supervised Semantic Segmentation. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)* (2021), 6944–6953.
[274] Yujia Li, Kevin Swersky, and Richard Zemel. 2015. Generative Moment Matching Networks. arXiv:cs.LG/1502.02761
[275] Y. Li, Tao Wang, Bingyi Kang, Sheng Tang, C. Wang, J. Li, and Jiashi Feng. 2020. Overcoming Classifier Imbalance for Long-tail Object Detection with Balanced Group Softmax. *ArXiv* abs/2006.10408 (2020).
[276] Yuanzhi Li and Yang Yuan. 2017. Convergence Analysis of Two-layer Neural Networks with ReLU Activation. In *NIPS*.
[277] Zeming Li, Chao Peng, Gang Yu, Xiangyu Zhang, Yangdong Deng, and Jian Sun. 2018. Detnet: A backbone network for object detection. *arXiv preprint arXiv:1804.06215* (2018).
[278] Shiyu Liang, Yixuan Li, and R. Srikant. 2018. Enhancing The Reliability of Out-of-distribution Image Detection in Neural Networks. *arXiv: Learning* (2018).
[279] Ming Liao. 2013. Applied Stochastic Processes.
[280] Qianli Liao and Tomaso A. Poggio. 2016. Bridging the Gaps Between Residual Learning, Recurrent Neural Networks and Visual Cortex. *ArXiv* abs/1604.03640 (2016).
[281] Sungbin Lim, Ildoo Kim, Taesup Kim, Chiheon Kim, and Sungwoong Kim. 2019. Fast autoaugment. In *Advances in Neural Information Processing Systems*. 6665–6675.
[282] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. 2017. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2117–2125.
[283] Tsung-Yi Lin, Piotr Dollár, Ross B. Girshick, Kaiming He, Bharath Hariharan, and Serge J. Belongie. 2017. Feature Pyramid Networks for Object Detection. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017), 936–944.
[284] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. 2017. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*. 2980–2988.
[285] Tsung-Yu Lin, Aruni RoyChowdhury, and Subhransu Maji. 2015. Bilinear CNNs for Fine-grained Visual Recognition. *arXiv: Computer Vision and Pattern Recognition* (2015).
[286] Fayao Liu, Chunhua Shen, Guosheng Lin, and Ian D. Reid. 2016. Learning Depth from Single Monocular Images Using Deep Convolutional Neural Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 38 (2016), 2024–2039.
[287] Hanxiao Liu, Andrew Brock, Karen Simonyan, and Quoc V. Le. 2020. Evolving Normalization-Activation Layers. *ArXiv* abs/2004.02967 (2020).
[288] Li Liu, Wanli Ouyang, Xiaogang Wang, Paul Fieguth, Jie Chen, Xinwang Liu, and Matti Pietikäinen. 2020. Deep learning for generic object detection: A survey. *International journal of computer vision* 128, 2 (2020), 261–318.
[289] Songtao Liu, Di Huang, et al. 2018. Receptive field block net for accurate and fast object detection. In *Proceedings of the European Conference on Computer Vision (ECCV)*. 385–400.
[290] Songtao Liu, Di Huang, and Yunhong Wang. 2018. Receptive Field Block Net for Accurate and Fast Object Detection. *ArXiv* abs/1711.07767 (2018).
[291] Shu Liu, Lu Qi, Haifang Qin, Jianping Shi, and Jiaya Jia. 2018. Path Aggregation Network for Instance Segmentation. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2018), 8759–8768.
[292] Shu Liu, Lu Qi, Haifang Qin, Jianping Shi, and Jiaya Jia. 2018. Path aggregation network for instance segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 8759–8768.
[293] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. 2016. Ssd: Single shot multibox detector. In *European conference on computer vision*. Springer, 21–37.
[294] Yen-Cheng Liu, Chih-Yao Ma, Zijian He, Chia-Wen Kuo, Kan Chen, Peizhao Zhang, Bichen Wu, Zsolt Kira, and Péter Vajda. 2021. Unbiased Teacher for Semi-Supervised Object Detection. *ArXiv* abs/2102.09480 (2021).
[295] Zhuang Liu, Hanzi Mao, Chaozheng Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. 2022. A ConvNet for the 2020s.
[296] Philipp Liznerski, Lukas Ruff, Robert A. Vandermeulen, Billy Joe Franks, M. Kloft, and Klaus-Robert Muller. 2021. Explainable Deep One-Class Classification. *ArXiv* abs/2007.01760 (2021).
[297] Jonathan Long, Evan Shelhamer, and Trevor Darrell. 2015. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3431–3440.

[298] Xiang Long, Kaipeng Deng, Guanzhong Wang, Yang Zhang, Qingqing Dang, Yuan Gao, Hui Shen, Jianguo Ren, Shumin Han, Errui Ding, et al. 2020. PP-YOLO: An Effective and Efficient Implementation of Object Detector. *arXiv preprint arXiv:2007.12099* (2020).
[299] Raphael Gontijo Lopes, Dong Yin, Ben Poole, Justin Gilmer, and Ekin Dogus Cubuk. 2019. Improving Robustness Without Sacrificing Accuracy with Patch Gaussian Augmentation. *ArXiv* abs/1906.02611 (2019).
[300] Ilya Loshchilov and Frank Hutter. 2017. SGDR: Stochastic Gradient Descent with Warm Restarts. In *ICLR*.
[301] Ilya Loshchilov and Frank Hutter. 2019. Decoupled Weight Decay Regularization. In *ICLR*.
[302] Ali Lotfi-Rezaabad and S. Vishwanath. 2020. Learning Representations by Maximizing Mutual Information in Variational Autoencoders. *2020 IEEE International Symposium on Information Theory (ISIT)* (2020), 2729–2734.
[303] G LoweDavid. 2004. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision* (2004).
[304] Zhou Lu, Hongming Pu, Feicheng Wang, Zhiqiang Hu, and Liwei Wang. 2017. The Expressive Power of Neural Networks: A View from the Width. In *NIPS*.
[305] W. Luo, Y. Li, R. Urtasun, and R. Zemel. 2016. Understanding the Effective Receptive Field in Deep Convolutional Neural Networks. *ArXiv* abs/1701.04128 (2016).
[306] Ningning Ma, Xiangyu Zhang, Haitao Zheng, and Jian Sun. 2018. ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design. *ArXiv* abs/1807.11164 (2018).
[307] X. Ma, Chunting Zhou, and E. Hovy. 2019. MAE: Mutual Posterior-Divergence Regularization for Variational AutoEncoders. *ArXiv* abs/1901.01498 (2019).
[308] Y. Ma, Stefano Soatto, Jana Koseck, and S. Sastry. 2004. An Invitation to 3-D Vision: From Images to Geometric Models.
[309] Subhransu Maji, Esa Rahtu, Juho Kannala, Matthew Blaschko, and Andrea Vedaldi. 2013. Fine-grained visual classification of aircraft. *arXiv preprint arXiv:1306.5151* (2013).
[310] Puneet Mangla, Nupur Kumari, Abhishek Sinha, Mayank Singh, Balaji Krishnamurthy, and Vineeth N Balasubramanian. 2020. Charting the right manifold: Manifold mixup for few-shot learning. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 2218–2227.
[311] Xudong Mao, Qing Li, Haoran Xie, Raymond Y. K. Lau, Zhen Wang, and Stephen Paul Smolley. 2016. Least Squares Generative Adversarial Networks. arXiv:cs.CV/1611.04076
[312] Xudong Mao, Qing Li, Haoran Xie, Raymond Y. K. Lau, Zhen Wang, and Stephen Paul Smolley. 2017. Least Squares Generative Adversarial Networks. *2017 IEEE International Conference on Computer Vision (ICCV)* (2017), 2813–2821.
[313] James Martens. 2016. Second-order Optimization for Neural Networks.
[314] P. McCullagh and J. Nelder. 1972. Generalized Linear Models.
[315] Sachin Mehta and Mohammad Rastegari. 2021. MobileViT: Light-weight, General-purpose, and Mobile-friendly Vision Transformer. *ArXiv* abs/2110.02178 (2021).
[316] Iaroslav Melekhov, Juho Kannala, and Esa Rahtu. 2017. Image Patch Matching Using Convolutional Descriptors with Euclidean Distance. arXiv:cs.CV/1710.11359
[317] Luke Metz, Ben Poole, David Pfau, and Jascha Sohl-Dickstein. 2016. Unrolled Generative Adversarial Networks. arXiv:cs.LG/1611.02163
[318] Hrushikesh Mhaskar, Qianli Liao, and Tomaso A. Poggio. 2017. When and Why Are Deep Networks Better Than Shallow Ones?. In *AAAI*.
[319] Mehdi Mirza and Simon Osindero. 2014. Conditional Generative Adversarial Nets. arXiv:cs.LG/1411.1784
[320] Anastasiya Mishchuk, Dmytro Mishkin, Filip Radenovic, and Jiri Matas. 2017. Working hard to know your neighbor's margins: Local descriptor learning loss. arXiv:cs.CV/1705.10872
[321] Dmytro Mishkin and Jiri Matas. 2016. All you need is a good init. *CoRR* abs/1511.06422 (2016).
[322] Diganta Misra. 2019. Mish: A self regularized non-monotonic neural activation function. *arXiv preprint arXiv:1908.08681* (2019).
[323] Ishan Misra and Laurens van der Maaten. 2020. Self-Supervised Learning of Pretext-Invariant Representations. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2020), 6706–6716.
[324] Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida. 2018. Spectral Normalization for Generative Adversarial Networks. arXiv:cs.LG/1802.05957
[325] Takeru Miyato and Masanori Koyama. 2018. cGANs with Projection Discriminator. arXiv:cs.LG/1802.05637
[326] Takeru Miyato, Shin-ichi Maeda, Masanori Koyama, and Shin Ishii. 2018. Virtual adversarial training: a regularization method for supervised and semi-supervised learning. *IEEE transactions on pattern analysis and machine intelligence* 41, 8 (2018), 1979–1993.
[327] Shakir Mohamed and Balaji Lakshminarayanan. 2016. Learning in Implicit Generative Models. arXiv:stat.ML/1610.03483
[328] Grégoire Montavon, Sebastian Lapuschkin, Alexander Binder, Wojciech Samek, and Klaus-Robert Müller. 2017. Explaining nonlinear classification decisions with deep Taylor decomposition. *Pattern Recognit.* 65 (2017), 211–222.
[329] Guido Montúfar, Razvan Pascanu, Kyunghyun Cho, and Yoshua Bengio. 2014. On the Number of Linear Regions of Deep Neural Networks. *ArXiv* abs/1402.1869 (2014).
[330] Roozbeh Mottaghi, Xianjie Chen, Xiaobai Liu, Nam-Gyu Cho, Seong-Whan Lee, Sanja Fidler, Raquel Urtasun, and Alan Yuille. 2014. The role of context for object detection and semantic segmentation in the wild. In *Proceedings of the IEEE Conference on Computer*

Vision and Pattern Recognition. 891–898.

[331] Yair Movshovitz-Attias, Alexander Toshev, Thomas Leung, Sergey Ioffe, and Saurabh Singh. 2017. No Fuss Distance Metric Learning Using Proxies. *2017 IEEE International Conference on Computer Vision (ICCV)* (2017), 360–368.

[332] Nairouz Mrabah, Naimul Mefraz Khan, Riadh Ksantini, and Zied Lachiri. 2020. Deep clustering with a dynamic autoencoder: From reconstruction towards centroids construction. *Neural Networks* 130 (2020), 206–228.

[333] Krikamol Muandet, Kenji Fukumizu, Bharath K. Sriperumbudur, and Bernhard Schölkopf. 2017. Kernel Mean Embedding of Distributions: A Review and Beyonds. *ArXiv* abs/1605.09522 (2017).

[334] Sudipto Mukherjee, Himanshu Asnani, Eugene Lin, and Sreeram Kannan. 2019. Clustergan: Latent space clustering in generative adversarial networks. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 33. 4610–4617.

[335] Jishnu Mukhoti, Viveka Kulharia, Amartya Sanyal, Stuart Golodetz, Philip Torr, and Puneet Dokania. 2020. Calibrating deep neural networks using focal loss. *Advances in Neural Information Processing Systems* 33 (2020), 15288–15299.

[336] Mahesh Chandra Mukkamala and Matthias Hein. 2017. Variants of RMSProp and Adagrad with Logarithmic Regret Bounds. In *ICML*.

[337] Rafael Müller, Simon Kornblith, and Geoffrey E. Hinton. 2019. When Does Label Smoothing Help?. In *NeurIPS*.

[338] Samuel Müller and Frank Hutter. 2021. TrivialAugment: Tuning-free Yet State-of-the-Art Data Augmentation. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)* (2021), 754–762.

[339] Terrell N. Mundhenk, Daniel Ho, and Barry Y. Chen. 2018. Improvements to Context Based Self-Supervised Learning. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2018), 9339–9348.

[340] Kevin P. Murphy. 2012. Machine learning - a probabilistic perspective. In *Adaptive computation and machine learning series*.

[341] Kimia Nadjahi, Alain Durmus, Umut Simsekli, and R. Badeau. 2019. Asymptotic Guarantees for Learning Generative Models with the Sliced-Wasserstein Distance. In *NeurIPS*.

[342] Brady Neal, Sarthak Mittal, A. Baratin, Vinayak Tantia, Matthew Scicluna, S. Lacoste-Julien, and Ioannis Mitliagkas. 2018. A Modern Take on the Bias-Variance Tradeoff in Neural Networks. *ArXiv* abs/1810.08591 (2018).

[343] Behnam Neyshabur, Hanie Sedghi, and Chiyuan Zhang. 2020. What is being transferred in transfer learning? *ArXiv* abs/2008.11687 (2020).

[344] Behnam Neyshabur, Hanie Sedghi, and Chiyuan Zhang. 2020. What is being transferred in transfer learning? *Advances in neural information processing systems* 33 (2020), 512–523.

[345] Quynh Nguyen and Matthias Hein. 2017. The Loss Surface of Deep and Wide Neural Networks. *ArXiv* abs/1704.08045 (2017).

[346] Tu Dinh Nguyen, Trung Le, Hung Vu, and Dinh Phung. 2017. Dual Discriminator Generative Adversarial Nets. arXiv:cs.LG/1709.03831

[347] Chuang Niu and Ge Wang. 2021. SPICE: Semantic Pseudo-labeling for Image Clustering. *ArXiv* abs/2103.09382 (2021).

[348] Arild Nøkland and Lars Hiller Eidnes. 2019. Training Neural Networks with Local Error Signals. In *ICML*.

[349] Sebastian Nowozin, Botond Cseke, and Ryota Tomioka. 2016. f-GAN: Training Generative Neural Samplers using Variational Divergence Minimization. arXiv:stat.ML/1606.00709

[350] Farzan Erlik Nowruzi, Prince Kapoor, Dhanvin Kolhatkar, Fahed Al Hassanat, Robert Laganiere, and Julien Rebut. 2019. How much real data do we actually need: Analyzing object detection performance using synthetic and real data. arXiv:cs.CV/1907.07061

[351] Augustus Odena. 2016. Semi-Supervised Learning with Generative Adversarial Networks. arXiv:stat.ML/1606.01583

[352] Augustus Odena, Christopher Olah, and Jonathon Shlens. 2016. Conditional Image Synthesis With Auxiliary Classifier GANs. arXiv:stat.ML/1610.09585

[353] Kemal Oksuz, Baris Can Cam, Sinan Kalkan, and Emre Akbas. 2021. Imbalance Problems in Object Detection: A Review. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43 (2021), 3388–3415.

[354] O. Oktay, Jo Schlemper, Loïc Le Folgoc, M. J. Lee, M. Heinrich, K. Misawa, K. Mori, Steven G. McDonagh, N. Hammerla, Bernhard Kainz, Ben Glocker, and D. Rueckert. 2018. Attention U-Net: Learning Where to Look for the Pancreas. *ArXiv* abs/1804.03999 (2018).

[355] Avital Oliver, Augustus Odena, Colin A Raffel, Ekin Dogus Cubuk, and Ian Goodfellow. 2018. Realistic evaluation of deep semi-supervised learning algorithms. *Advances in neural information processing systems* 31 (2018).

[356] Sung Woo Park and Junseok Kwon. 2019. Sphere Generative Adversarial Network Based on Geometric Moment Matching. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2019), 4287–4296.

[357] Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. 2019. Semantic Image Synthesis with Spatially-Adaptive Normalization. arXiv:cs.CV/1903.07291

[358] Jack Parker-Holder, Vu Nguyen, and Stephen J. Roberts. 2020. Provably Efficient Online Hyperparameter Optimization with Population-Based Bandits. *arXiv: Learning* (2020).

[359] Deepak Pathak, Philipp Krahenbuhl, Jeff Donahue, Trevor Darrell, and Alexei A. Efros. 2016. Context Encoders: Feature Learning by Inpainting. arXiv:cs.CV/1604.07379

[360] Xingchao Peng, Ben Usman, Kuniaki Saito, Neela Kaushik, Judy Hoffman, and Kate Saenko. 2018. Syn2real: A new benchmark forsynthetic-to-real visual domain adaptation. *arXiv preprint arXiv:1806.09755* (2018).

[361] Federico Perazzi, Philipp Krähenbühl, Yael Pritch, and Alexander Sorkine-Hornung. 2012. Saliency filters: Contrast based filtering for salient region detection. *2012 IEEE Conference on Computer Vision and Pattern Recognition* (2012), 733–740.

[362] Pramuditha Perera, Ramesh Nallapati, and Bing Xiang. 2019. OCGAN: One-Class Novelty Detection Using GANs With Constrained Latent Representations. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2019), 2893–2901.
[363] Mohammad Pezeshki, Sekouba Kaba, Yoshua Bengio, Aaron C. Courville, Doina Precup, and Guillaume Lajoie. 2020. Gradient Starvation: A Learning Proclivity in Neural Networks. *ArXiv* abs/2011.09468 (2020).
[364] Hieu Pham, Melody Y. Guan, Barret Zoph, Quoc V. Le, and Jeff Dean. 2018. Efficient Neural Architecture Search via Parameter Sharing. In *ICML*.
[365] Hieu Pham, Qizhe Xie, Zihang Dai, and Quoc V. Le. 2021. Meta Pseudo Labels. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2021), 11552–11563.
[366] Pedro H. O. Pinheiro, Ronan Collobert, and Piotr Dollár. 2015. Learning to Segment Object Candidates. In *NIPS*.
[367] David C. Plaut and Martha J. Farah. 1990. Visual Object Representation: Interpreting Neurophysiological Data within a Computational Framework. *Journal of Cognitive Neuroscience* 2 (1990), 320–343.
[368] Guo-Jun Qi. 2017. Loss-Sensitive Generative Adversarial Networks on Lipschitz Densities. arXiv:cs.CV/1701.06264
[369] Alec Radford, Luke Metz, and Soumith Chintala. 2015. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. arXiv:cs.LG/1511.06434
[370] Ilija Radosavovic, Justin Johnson, Saining Xie, Wan-Yen Lo, and Piotr Dollár. 2019. On Network Design Spaces for Visual Recognition. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)* (2019), 1882–1890.
[371] Ilija Radosavovic, Raj Prateek Kosaraju, Ross B. Girshick, Kaiming He, and Piotr Dollár. 2020. Designing Network Design Spaces. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2020), 10425–10433.
[372] Aditi Raghunathan, Sang Michael Xie, Fanny Yang, John C. Duchi, and Percy Liang. 2019. Adversarial Training Can Hurt Generalization. *ArXiv* abs/1906.06032 (2019).
[373] Prajit Ramachandran, Barret Zoph, and Quoc V. Le. 2018. Searching for Activation Functions. *ArXiv* abs/1710.05941 (2018).
[374] René Ranftl, Alexey Bochkovskiy, and Vladlen Koltun. 2021. Vision Transformers for Dense Prediction. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)* (2021), 12159–12168.
[375] E. Real, A. Aggarwal, Y. Huang, and Quoc V. Le. 2019. Regularized Evolution for Image Classifier Architecture Search. *ArXiv* abs/1802.01548 (2019).
[376] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. 2016. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 779–788.
[377] Joseph Redmon and Ali Farhadi. 2017. YOLO9000: better, faster, stronger. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 7263–7271.
[378] Joseph Redmon and Ali Farhadi. 2018. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767* (2018).
[379] James K. Reed, Zach DeVito, Horace He, Ansley Ussery, and Jason Ansel. 2021. torch.fx: Practical Program Capture and Transformation for Deep Learning in Python. *ArXiv* abs/2112.08429 (2021).
[380] Scott Reed, Zeynep Akata, Xinchen Yan, Lajanugen Logeswaran, Bernt Schiele, and Honglak Lee. 2016. Generative Adversarial Text to Image Synthesis. arXiv:cs.NE/1605.05396
[381] Mark Reid, Robert Williamson, et al. 2011. Information, divergence and risk for binary experiments. (2011).
[382] Tal Remez, Jonathan Huang, and Matthew A. Brown. 2018. Learning to Segment via Cut-and-Paste. *ArXiv* abs/1803.06414 (2018).
[383] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*. 91–99.
[384] Shaoqing Ren, Kaiming He, Ross B. Girshick, and Jian Sun. 2015. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39 (2015), 1137–1149.
[385] Zhongzheng Ren, Zhiding Yu, Xiaodong Yang, Ming-Yu Liu, Yong Jae Lee, Alexander G. Schwing, and Jan Kautz. 2020. Instance-Aware, Context-Focused, and Memory-Efficient Weakly Supervised Object Detection. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2020), 10595–10604.
[386] Ives Rey-Otero and Mauricio Delbracio. 2014. Anatomy of the SIFT Method. *IPOL Journal* 4 (2014), 370–396.
[387] Hamid Rezatofighi, Nathan Tsoi, JunYoung Gwak, Amir Sadeghian, Ian Reid, and Silvio Savarese. 2019. Generalized intersection over union: A metric and a loss for bounding box regression. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 658–666.
[388] Danilo Jimenez Rezende, S. Mohamed, and Daan Wierstra. 2014. Stochastic Backpropagation and Approximate Inference in Deep Generative Models. In *ICML*.
[389] Christian P. Robert and George Casella. 2013. Introducing Monte Carlo methods with.
[390] Adriana Romero, Nicolas Ballas, Samira Ebrahimi Kahou, Antoine Chassang, Carlo Gatta, and Yoshua Bengio. 2014. FitNets: Hints for Thin Deep Nets. arXiv:cs.LG/1412.6550
[391] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*. Springer, 234–241.

[392] Abhijit Guha Roy, Nassir Navab, and Christian Wachinger. 2018. Concurrent Spatial and Channel Squeeze & Excitation in Fully Convolutional Networks. *ArXiv* abs/1803.02579 (2018).
[393] Noam Rozen, Aditya Grover, Maximilian Nickel, and Yaron Lipman. 2021. Moser Flow: Divergence-based Generative Modeling on Manifolds. *ArXiv* abs/2108.08052 (2021).
[394] Jose C Rubio, Joan Serrat, Antonio López, and Nikos Paragios. 2012. Unsupervised co-segmentation through region matching. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 749–756.
[395] Lukas Ruff, Nico Görnitz, Lucas Deecke, Shoaib Ahmed Siddiqui, Robert A. Vandermeulen, Alexander Binder, Emmanuel Müller, and M. Kloft. 2018. Deep One-Class Classification. In *ICML*.
[396] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. 2015. Imagenet large scale visual recognition challenge. *International journal of computer vision* 115, 3 (2015), 211–252.
[397] Kuniaki Saito, Yoshitaka Ushiku, Tatsuya Harada, and Kate Saenko. 2019. Strong-weak distribution alignment for adaptive object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 6956–6965.
[398] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. 2016. Improved Techniques for Training GANs. arXiv:cs.LG/1606.03498
[399] Tim Salimans and Diederik P. Kingma. 2016. Weight Normalization: A Simple Reparameterization to Accelerate Training of Deep Neural Networks. In *NIPS*.
[400] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. 2018. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4510–4520.
[401] Pedro Savarese and Michael Maire. 2019. Learning Implicitly Recurrent CNNs Through Parameter Sharing. arXiv:cs.LG/1902.09701
[402] Andrew M. Saxe, James L. McClelland, and Surya Ganguli. 2014. Exact solutions to the nonlinear dynamics of learning in deep linear neural networks. *CoRR* abs/1312.6120 (2014).
[403] Shreyas Saxena and Jakob Verbeek. 2016. Convolutional Neural Fabrics. *ArXiv* abs/1606.02492 (2016).
[404] Mischa Schmidt, Shahd Safarani, J. Gastinger, Tobias Jacobs, Sebastien Nicolas, and Anett Schülke. 2019. On the Performance of Differential Evolution for Hyperparameter Tuning. *2019 International Joint Conference on Neural Networks (IJCNN)* (2019), 1–8.
[405] Florian Schroff, Dmitry Kalenichenko, and James Philbin. 2015. FaceNet: A unified embedding for face recognition and clustering. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2015), 815–823.
[406] Ari Seff, Alex Beatson, Daniel Suo, and Han Liu. 2017. Continual Learning in Generative Adversarial Nets. arXiv:cs.LG/1705.08395
[407] Thomas Serre, Lior Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio. 2007. Robust Object Recognition with Cortex-Like Mechanisms. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29 (2007), 411–426.
[408] Thomas Serre, Lior Wolf, and T. Poggio. 2005. Object recognition with features inspired by visual cortex. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)* 2 (2005), 994–1000 vol. 2.
[409] Ali Shafahi, Mahyar Najibi, Amin Ghiasi, Zheng Xu, John P. Dickerson, Christoph Studer, Larry S. Davis, Gavin Taylor, and Tom Goldstein. 2019. Adversarial Training for Free!. In *NeurIPS*.
[410] Jian Shen, Yanru Qu, Weinan Zhang, and Yong Yu. 2017. Wasserstein Distance Guided Representation Learning for Domain Adaptation. arXiv:stat.ML/1707.01217
[411] Zhiqiang Shen, Zechun Liu, Zhuang Liu, Marios Savvides, Trevor Darrell, and Eric Xing. 2020. Un-mix: Rethinking image mixtures for unsupervised visual representation learning. *arXiv preprint arXiv:2003.05438* (2020).
[412] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P. Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. 2016. Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016), 1874–1883.
[413] Connor Shorten and Taghi M. Khoshgoftaar. 2019. A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data* 6 (2019), 1–48.
[414] Shai Shalev Shwartz and Shai Ben David. 2015. Understanding Machine Learning: From Theory To Algorithms.
[415] Patrice Y. Simard, David Steinkraus, and John C. Platt. 2003. Best practices for convolutional neural networks applied to visual document analysis. *Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings.* (2003), 958–963.
[416] Martin Simony, Stefan Milzy, Karl Amendey, and Horst-Michael Gross. 2018. Complex-yolo: An euler-region-proposal for real-time 3d object detection on point clouds. In *Proceedings of the European Conference on Computer Vision (ECCV)*. 0–0.
[417] Karen Simonyan and Andrew Zisserman. 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv:cs.CV/1409.1556
[418] Krishna Kumar Singh, Utkarsh Ojha, and Y. Lee. 2019. FineGAN: Unsupervised Hierarchical Disentanglement for Fine-Grained Object Generation and Discovery. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2019), 6483–6492.
[419] S. Singh, Derek Hoiem, and D. Forsyth. 2016. Swapout: Learning an ensemble of deep architectures. *ArXiv* abs/1605.06465 (2016).
[420] Saurabh Singh and Shankar Krishnan. 2020. Filter Response Normalization Layer: Eliminating Batch Dependence in the Training of Deep Neural Networks. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2020), 11234–11243.

[421] Leslie N. Smith. 2017. Cyclical Learning Rates for Training Neural Networks. *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)* (2017), 464–472.

[422] Leslie N. Smith. 2018. A disciplined approach to neural network hyper-parameters: Part 1 - learning rate, batch size, momentum, and weight decay. *ArXiv* abs/1803.09820 (2018).

[423] Leslie N Smith. 2018. A disciplined approach to neural network hyper-parameters: Part 1–learning rate, batch size, momentum, and weight decay. *arXiv preprint arXiv:1803.09820* (2018).

[424] Samuel L. Smith, Pieter-Jan Kindermans, and Quoc V. Le. 2018. Don't Decay the Learning Rate, Increase the Batch Size. *ArXiv* abs/1711.00489 (2018).

[425] Konstantin Sofiiuk, Olga Barinova, and Anton Konushin. 2019. Adaptis: Adaptive instance selection network. In *Proceedings of the IEEE International Conference on Computer Vision*. 7355–7363.

[426] Kihyuk Sohn. 2016. Improved Deep Metric Learning with Multi-class N-pair Loss Objective. In *NIPS*.

[427] Kihyuk Sohn, David Berthelot, Nicholas Carlini, Zizhao Zhang, Han Zhang, Colin A Raffel, Ekin Dogus Cubuk, Alexey Kurakin, and Chun-Liang Li. 2020. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *Advances in Neural Information Processing Systems* 33 (2020), 596–608.

[428] Jiaming Song and Stefano Ermon. 2019. Understanding the limitations of variational mutual information estimators. *arXiv preprint arXiv:1910.06222* (2019).

[429] Xinhang Song, Luis Herranz, and Shuqiang Jiang. 2017. Depth CNNs for RGB-D Scene Recognition: Learning from Scratch Better than Transferring from RGB-CNNs. In *AAAI*.

[430] Yang Song and Stefano Ermon. 2019. Generative Modeling by Estimating Gradients of the Data Distribution. arXiv:cs.LG/1907.05600

[431] G. Sperling. 1963. A Model for Visual Memory Tasks1. *Human Factors: The Journal of Human Factors and Ergonomics Society* 5 (1963), 19 – 31.

[432] Jost Tobias Springenberg, Alexey Dosovitskiy, Thomas Brox, and Martin Riedmiller. 2014. Striving for Simplicity: The All Convolutional Net. arXiv:cs.LG/1412.6806

[433] Suvrit Sra, Sebastian Nowozin, and Stephen J. Wright. 2011. *Optimization for Machine Learning*. The MIT Press.

[434] A. Srinivas, Tsung-Yi Lin, Niki Parmar, Jonathon Shlens, P. Abbeel, and Ashish Vaswani. 2021. Bottleneck Transformers for Visual Recognition. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2021), 16514–16524.

[435] Bharath K. Sriperumbudur, Kenji Fukumizu, Arthur Gretton, Bernhard Schölkopf, and Gert R. G. Lanckriet. 2012. On the empirical estimation of integral probability metrics. *Electronic Journal of Statistics* 6 (2012), 1550–1599.

[436] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research* 15, 1 (2014), 1929–1958.

[437] Rupesh Kumar Srivastava, Klaus Greff, and Jürgen Schmidhuber. 2015. Highway Networks. arXiv:cs.LG/1505.00387

[438] Rainer Storn and Kenneth V. Price. 1997. Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces. *Journal of Global Optimization* 11 (1997), 341–359.

[439] Weijie Su, Stephen P. Boyd, and Emmanuel J. Candès. 2014. A Differential Equation for Modeling Nesterov's Accelerated Gradient Method: Theory and Insights. In *J. Mach. Learn. Res.*

[440] Z. Su, Linpu Fang, Wenxiong Kang, Dewen Hu, Matti Pietikäinen, and Li Liu. 2020. Dynamic Group Convolution for Accelerating Convolutional Neural Networks. In *ECCV*.

[441] Sainbayar Sukhbaatar, Joan Bruna, Manohar Paluri, Lubomir Bourdev, and Rob Fergus. 2014. Training convolutional networks with noisy labels. *arXiv preprint arXiv:1406.2080* (2014).

[442] Jiaming Sun, Zehong Shen, Yuang Wang, Hujun Bao, and Xiaowei Zhou. 2021. LoFTR: Detector-Free Local Feature Matching with Transformers. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2021), 8918–8927.

[443] Ilya Sutskever, James Martens, George E. Dahl, and Geoffrey E. Hinton. 2013. On the importance of initialization and momentum in deep learning. In *ICML*.

[444] Taiji Suzuki and Atsushi Nitanda. 2019. Deep learning is adaptive to intrinsic dimensionality of model smoothness in anisotropic Besov space. *ArXiv* abs/1910.12799 (2019).

[445] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alex Alemi. 2016. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. arXiv:cs.CV/1602.07261

[446] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. 2014. Going Deeper with Convolutions. arXiv:cs.CV/1409.4842

[447] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna. 2015. Rethinking the Inception Architecture for Computer Vision. arXiv:cs.CV/1512.00567

[448] R. Szeliski. 2011. Computer Vision - Algorithms and Applications. In *Texts in Computer Science*.

[449] Mingxing Tan and Quoc Le. 2021. Efficientnetv2: Smaller models and faster training. In *International Conference on Machine Learning*. PMLR, 10096–10106.

[450] Mingxing Tan and Quoc V. Le. 2019. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. arXiv:cs.LG/1905.11946

[451] Mingxing Tan and Quoc V. Le. 2019. MixConv: Mixed Depthwise Convolutional Kernels. *ArXiv* abs/1907.09595 (2019).
[452] Mingxing Tan, Ruoming Pang, and Quoc V. Le. 2020. EfficientDet: Scalable and Efficient Object Detection. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2020), 10778–10787.
[453] Mingxing Tan, Ruoming Pang, and Quoc V Le. 2020. Efficientdet: Scalable and efficient object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 10781–10790.
[454] Matthew Tancik, Pratul P. Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan T. Barron, and Ren Ng. 2020. Fourier Features Let Networks Learn High Frequency Functions in Low Dimensional Domains. *ArXiv* abs/2006.10739 (2020).
[455] Peng Tang, Xinggang Wang, Angtian Wang, Yongluan Yan, Wenyu Liu, Junzhou Huang, and Alan Loddon Yuille. 2018. Weakly Supervised Region Proposal Network and Object Detection. In *ECCV*.
[456] Sasha Targ, D. Almeida, and Kevin Lyman. 2016. Resnet in Resnet: Generalizing Residual Architectures. *ArXiv* abs/1603.08029 (2016).
[457] Matus Telgarsky. 2016. Benefits of Depth in Neural Networks. *ArXiv* abs/1602.04485 (2016).
[458] Martin Thoma. 2016. A Survey of Semantic Segmentation. *ArXiv* abs/1602.06541 (2016).
[459] Sunil Thulasidasan, Gopinath Chennupati, Jeff A. Bilmes, Tanmoy Bhattacharya, and Sarah Ellen Michalak. 2019. On Mixup Training: Improved Calibration and Predictive Uncertainty for Deep Neural Networks. In *NeurIPS*.
[460] Yonglong Tian, Chen Sun, Ben Poole, Dilip Krishnan, Cordelia Schmid, and Phillip Isola. 2020. What makes for good views for contrastive learning. *ArXiv* abs/2005.10243 (2020).
[461] Zhi Tian, Chunhua Shen, Hao Chen, and Tong He. 2019. FCOS: Fully Convolutional One-Stage Object Detection. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)* (2019), 9626–9635.
[462] Zhi Tian, Chunhua Shen, Hao Chen, and Tong He. 2019. Fcos: Fully convolutional one-stage object detection. In *Proceedings of the IEEE international conference on computer vision*. 9627–9636.
[463] Zhi Tian, Chunhua Shen, Xinlong Wang, and Hao Chen. 2021. BoxInst: High-Performance Instance Segmentation with Box Annotations. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2021), 5439–5448.
[464] Ilya O. Tolstikhin, Neil Houlsby, Alexander Kolesnikov, Lucas Beyer, Xiaohua Zhai, Thomas Unterthiner, Jessica Yung, Daniel Keysers, Jakob Uszkoreit, Mario Lucic, and Alexey Dosovitskiy. 2021. MLP-Mixer: An all-MLP Architecture for Vision. *ArXiv* abs/2105.01601 (2021).
[465] Ilya O. Tolstikhin, Bharath K. Sriperumbudur, and Bernhard Schölkopf. 2016. Minimax Estimation of Maximum Mean Discrepancy with Radial Kernels. In *NIPS*.
[466] Antonio Torralba and Alexei A. Efros. 2011. Unbiased look at dataset bias. *CVPR 2011* (2011), 1521–1528.
[467] Florian Tramèr, Alexey Kurakin, Nicolas Papernot, Dan Boneh, and Patrick Mcdaniel. 2018. Ensemble Adversarial Training: Attacks and Defenses. *ArXiv* abs/1705.07204 (2018).
[468] Michael Tschannen, Olivier Bachem, and M. Lucic. 2018. Recent Advances in Autoencoder-Based Representation Learning. *ArXiv* abs/1812.05069 (2018).
[469] Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. 2017. Adversarial discriminative domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 7167–7176.
[470] Eric Tzeng, Judy Hoffman, Ning Zhang, Kate Saenko, and Trevor Darrell. 2014. Deep domain confusion: Maximizing for domain invariance. *arXiv preprint arXiv:1412.3474* (2014).
[471] Matej Ulicny, Vladimir A. Krylov, and Rozenn Dahyot. 2019. Harmonic Networks with Limited Training Samples. arXiv:cs.CV/1905.00135
[472] Aäron van den Oord, Yazhe Li, and Oriol Vinyals. 2018. Representation Learning with Contrastive Predictive Coding. *ArXiv* abs/1807.03748 (2018).
[473] Jesper E Van Engelen and Holger H Hoos. 2020. A survey on semi-supervised learning. *Machine Learning* 109, 2 (2020), 373–440.
[474] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, L. Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. *ArXiv* abs/1706.03762 (2017).
[475] Andreas Veit, Michael J. Wilber, and Serge J. Belongie. 2016. Residual Networks Behave Like Ensembles of Relatively Shallow Networks. In *NIPS*.
[476] Vikas Verma, Kenji Kawaguchi, Alex Lamb, Juho Kannala, Yoshua Bengio, and David Lopez-Paz. 2019. Interpolation consistency training for semi-supervised learning. *arXiv preprint arXiv:1903.03825* (2019).
[477] Vikas Verma, Alex Lamb, Christopher Beckham, Amir Najafi, Ioannis Mitliagkas, David Lopez-Paz, and Yoshua Bengio. 2019. Manifold mixup: Better representations by interpolating hidden states. In *International Conference on Machine Learning*. 6438–6447.
[478] Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol. 2008. Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th international conference on Machine learning*. 1096–1103.
[479] Catherine Wah, Steve Branson, Peter Welinder, Pietro Perona, and Serge Belongie. 2011. The caltech-ucsd birds-200-2011 dataset. (2011).
[480] Jacob Walker, Carl Doersch, Abhinav Gupta, and Martial Hebert. 2016. An uncertain future: Forecasting from static images using variational autoencoders. In *European Conference on Computer Vision*. Springer, 835–851.

[481] Bao Wang, Xiyang Luo, Zhen Li, Wei Zhu, Zuoqiang Shi, and Stanley Osher. 2018. Deep neural nets with interpolating function as output activation. *Advances in neural information processing systems* 31 (2018).
[482] C. Wang, K. L. Tan, and C. Lin. 2020. Newton Methods for Convolutional Neural Networks. *ACM Transactions on Intelligent Systems and Technology (TIST)* 11 (2020), 1 – 30.
[483] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. 2021. Scaled-YOLOv4: Scaling Cross Stage Partial Network. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2021), 13024–13033.
[484] Chien-Yao Wang, Hong-Yuan Mark Liao, Yueh-Hua Wu, Ping-Yang Chen, Jun-Wei Hsieh, and I-Hau Yeh. 2020. CSPNet: A new backbone that can enhance learning capability of CNN. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*. 390–391.
[485] Chien-Yao Wang, Hong-Yuan Mark Liao, I-Hau Yeh, Yueh-Hua Wu, Ping-Yang Chen, and Jun-Wei Hsieh. 2020. CSPNet: A New Backbone that can Enhance Learning Capability of CNN. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (2020), 1571–1580.
[486] Jing Wang, Jiahong Chen, Jianzhe Lin, Leonid Sigal, and Clarence W de Silva. 2020. Discriminative Feature Alignment: Improving Transferability of Unsupervised Domain Adaptation by Gaussian-guided Latent Alignment. *arXiv preprint arXiv:2006.12770* (2020).
[487] Jiaqi Wang, Kai Chen, Shuo Yang, Chen Change Loy, and Dahua Lin. 2019. Region Proposal by Guided Anchoring. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2019), 2960–2969.
[488] Jianan Wang, Eren Sezener, David Budden, Marcus Hutter, and Joel Veness. 2020. A combinatorial perspective on transfer learning. *Advances in Neural Information Processing Systems* 33 (2020), 918–929.
[489] Jingdong Wang, K. Sun, Tianheng Cheng, Borui Jiang, C. Deng, Y. Zhao, D. Liu, Y. Mu, Mingkui Tan, Xinggang Wang, W. Liu, and Bin Xiao. 2020. Deep High-Resolution Representation Learning for Visual Recognition. *IEEE transactions on pattern analysis and machine intelligence* (2020).
[490] Jiaqi Wang, Wenwei Zhang, Yuhang Cao, Kai Chen, Jiangmiao Pang, Tao Gong, Jianping Shi, Chen Change Loy, and Dahua Lin. 2020. Side-Aware Boundary Localization for More Precise Object Detection. *ArXiv* abs/1912.04260 (2020).
[491] Robert J Wang, Xiang Li, and Charles X Ling. 2018. Pelee: A real-time object detection system on mobile devices. *Advances in neural information processing systems* 31 (2018).
[492] Tongzhou Wang and Phillip Isola. 2020. Understanding Contrastive Representation Learning through Alignment and Uniformity on the Hypersphere. In *ICML*.
[493] Wenguan Wang, Qiuxia Lai, Huazhu Fu, Jianbing Shen, and Haibin Ling. 2019. Salient Object Detection in the Deep Learning Era: An In-Depth Survey. *ArXiv* abs/1904.09146 (2019).
[494] X. Wang, Ross B. Girshick, Abhinav Kumar Gupta, and Kaiming He. 2018. Non-local Neural Networks. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2018), 7794–7803.
[495] Xun Wang, Xintong Han, Weilin Huang, Dengke Dong, and Matthew R. Scott. 2019. Multi-Similarity Loss With General Pair Weighting for Deep Metric Learning. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2019), 5017–5025.
[496] Xinlong Wang, Tao Kong, Chunhua Shen, Yuning Jiang, and Lei Li. 2019. Solo: Segmenting objects by locations. *arXiv preprint arXiv:1912.04488* (2019).
[497] Xinlong Wang, Rufeng Zhang, Tao Kong, Lei Li, and Chunhua Shen. 2020. SOLOv2: Dynamic, Faster and Stronger. *arXiv preprint arXiv:2003.10152* (2020).
[498] Yunhe Wang, Chang Xu, Chao Xu, and Dacheng Tao. 2017. Beyond Filters: Compact Feature Map for Portable Deep Model. In *Proceedings of the 34th International Conference on Machine Learning (Proceedings of Machine Learning Research)*, Doina Precup and Yee Whye Teh (Eds.), Vol. 70. PMLR, International Convention Centre, Sydney, Australia, 3703–3711. http://proceedings.mlr.press/v70/wang17m.html
[499] Yude Wang, Jie Zhang, Meina Kan, S. Shan, and Xilin Chen. 2020. Self-Supervised Equivariant Attention Mechanism for Weakly Supervised Semantic Segmentation. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2020), 12272–12281.
[500] Yingming Wang, X. Zhang, Tong Yang, and Jian Sun. 2021. Anchor DETR: Query Design for Transformer-Based Detector. *ArXiv* abs/2109.07107 (2021).
[501] Zhengwei Wang, Qi She, and Tomas E. Ward. 2019. Generative Adversarial Networks in Computer Vision: A Survey and Taxonomy. arXiv:cs.LG/1906.01529
[502] Longhui Wei, An Xiao, Lingxi Xie, Xin Chen, Xiaopeng Zhang, and Qi Tian. 2020. Circumventing Outliers of AutoAugment with Knowledge Distillation. *arXiv preprint arXiv:2003.11342* (2020).
[503] Yunchao Wei, Jiashi Feng, Xiaodan Liang, Ming-Ming Cheng, Yao Zhao, and Shuicheng Yan. 2017. Object Region Mining with Adversarial Erasing: A Simple Classification to Semantic Segmentation Approach. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017), 6488–6496.
[504] Yunchao Wei, Huaxin Xiao, Honghui Shi, Zequn Jie, Jiashi Feng, and Thomas S Huang. 2018. Revisiting dilated convolution: A simple approach for weakly-and semi-supervised semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 7268–7277.
[505] Lilian Weng. 2019. From GAN to WGAN. arXiv:cs.LG/1904.08994

[506] M. Wheeler and A. Treisman. 2002. Binding in short-term visual memory. *Journal of experimental psychology. General* 131 1 (2002), 48–64.
[507] Thomas Wiatowski and H. Bölcskei. 2018. A Mathematical Theory of Deep Convolutional Neural Networks for Feature Extraction. *IEEE Transactions on Information Theory* 64 (2018), 1845–1866.
[508] Svante Wold, Kim H. Esbensen, and Paul Geladi. 1987. Principal component analysis. *Chemometrics and Intelligent Laboratory Systems* 2 (1987), 37–52.
[509] Sebastien C. Wong, Adam Gatt, Victor Stamatescu, and Mark D. McDonnell. 2016. Understanding Data Augmentation for Classification: When to Warp? *2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA)* (2016), 1–6.
[510] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In-So Kweon. 2018. CBAM: Convolutional Block Attention Module. In *ECCV*.
[511] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. 2018. CBAM: Convolutional Block Attention Module. arXiv:cs.CV/1807.06521
[512] Yuxin Wu and Kaiming He. 2018. Group normalization. In *Proceedings of the European conference on computer vision (ECCV)*. 3–19.
[513] Yuxin Wu and Justin Johnson. 2021. Rethinking "Batch" in BatchNorm. *ArXiv* abs/2105.07576 (2021).
[514] Zhirong Wu, Yuanjun Xiong, Stella X. Yu, and Dahua Lin. 2018. Unsupervised Feature Learning via Non-parametric Instance Discrimination. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2018), 3733–3742.
[515] Yongqin Xian, Christoph H. Lampert, Bernt Schiele, and Zeynep Akata. 2019. Zero-Shot Learning—A Comprehensive Evaluation of the Good, the Bad and the Ugly. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41 (2019), 2251–2265.
[516] Bo Xie, Yingyu Liang, and Le Song. 2017. Diverse Neural Network Learns True Target Functions. In *AISTATS*.
[517] Cihang Xie, Mingxing Tan, Boqing Gong, Jiang Wang, Alan Loddon Yuille, and Quoc V. Le. 2020. Adversarial Examples Improve Image Recognition. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2020), 816–825.
[518] Haozhe Xie, Hongxun Yao, Shengping Zhang, Shangchen Zhou, and Wenxiu Sun. 2020. Pix2Vox++: Multi-scale Context-aware 3D Object Reconstruction from Single and Multiple Images. *International Journal of Computer Vision* (2020), 1 – 17.
[519] Qizhe Xie, Zihang Dai, Eduard H. Hovy, Minh-Thang Luong, and Quoc V. Le. 2020. Unsupervised Data Augmentation for Consistency Training. *arXiv: Learning* (2020).
[520] Qizhe Xie, Eduard H. Hovy, Minh-Thang Luong, and Quoc V. Le. 2020. Self-Training With Noisy Student Improves ImageNet Classification. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2020), 10684–10695.
[521] Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. 2016. Aggregated Residual Transformations for Deep Neural Networks. arXiv:cs.CV/1611.05431
[522] Yuwen Xiong, Renjie Liao, Hengshuang Zhao, Rui Hu, Min Bai, Ersin Yumer, and Raquel Urtasun. 2019. Upsnet: A unified panoptic segmentation network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 8818–8826.
[523] Ismet Zeki Yalniz, Hervé Jégou, Kan Chen, Manohar Paluri, and Dhruv Kumar Mahajan. 2019. Billion-scale semi-supervised learning for image classification. *ArXiv* abs/1905.00546 (2019).
[524] Yoshihiro Yamada, Masakazu Iwamura, Takuya Akiba, and Koichi Kise. 2019. Shakedrop regularization for deep residual learning. *IEEE Access* 7 (2019), 186126–186136.
[525] Tong Yang, Xiangyu Zhang, Wenqiang Zhang, and Jian Sun. 2018. MetaAnchor: Learning to Detect Objects with Customized Anchors. In *NeurIPS*.
[526] Zili Yi, Hao Zhang, Ping Tan, and Minglun Gong. 2017. DualGAN: Unsupervised Dual Learning for Image-to-Image Translation. arXiv:cs.CV/1704.02510
[527] Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. 2014. How transferable are features in deep neural networks?. In *Advances in neural information processing systems*. 3320–3328.
[528] Yang You, Igor Gitman, and Boris Ginsburg. 2017. Large Batch Training of Convolutional Networks. *arXiv: Computer Vision and Pattern Recognition* (2017).
[529] B. Yu and Dacheng Tao. 2019. Deep Metric Learning With Tuplet Margin Loss. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)* (2019), 6489–6498.
[530] Fisher Yu, Dequan Wang, Evan Shelhamer, and Trevor Darrell. 2017. Deep Layer Aggregation. arXiv:cs.CV/1707.06484
[531] Lantao Yu, Weinan Zhang, Jun Wang, and Yong Yu. 2016. SeqGAN: Sequence Generative Adversarial Nets with Policy Gradient. arXiv:cs.LG/1609.05473
[532] Li Yuan, Qibin Hou, Zihang Jiang, Jiashi Feng, and Shuicheng Yan. 2021. VOLO: Vision Outlooker for Visual Recognition. *ArXiv* abs/2106.13112 (2021).
[533] Yuhui Yuan, Xilin Chen, and Jingdong Wang. 2019. Object-contextual representations for semantic segmentation. *arXiv preprint arXiv:1909.11065* (2019).
[534] Sangdoo Yun, Dongyoon Han, Seong Joon Oh, Sanghyuk Chun, Junsuk Choe, and Youngjoon Yoo. 2019. Cutmix: Regularization strategy to train strong classifiers with localizable features. In *Proceedings of the IEEE International Conference on Computer Vision*. 6023–6032.
[535] Sergey Zagoruyko and Nikos Komodakis. 2016. Wide Residual Networks. arXiv:cs.CV/1605.07146

[536] Arber Zela, Aaron Klein, Stefan Falkner, and Frank Hutter. 2018. Towards Automated Deep Learning: Efficient Joint Neural Architecture and Hyperparameter Search. *ArXiv* abs/1807.06906 (2018).

[537] Zhaoyang Zeng, Bei Liu, Jianlong Fu, Hongyang Chao, and Lei Zhang. 2019. WSOD2: Learning Bottom-Up and Top-Down Objectness Distillation for Weakly-Supervised Object Detection. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)* (2019), 8291–8299.

[538] Andrew Zhai and Hao-Yu Wu. 2019. Classification is a Strong Baseline for Deep Metric Learning. In *BMVC*.

[539] Shuangfei Zhai, Yu Cheng, Weining Lu, and Zhongfei Zhang. 2016. Doubly Convolutional Neural Networks. arXiv:cs.LG/1610.09716

[540] Xiaohua Zhai, Avital Oliver, Alexander Kolesnikov, and Lucas Beyer. 2019. S4L: Self-Supervised Semi-Supervised Learning. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)* (2019), 1476–1485.

[541] Xiaohua Zhai, Joan Puigcerver, Alexander Kolesnikov, Pierre Ruyssen, Carlos Riquelme, Mario Lucic, Josip Djolonga, Andre Susano Pinto, Maxim Neumann, Alexey Dosovitskiy, et al. 2019. A large-scale study of representation learning with the visual task adaptation benchmark. *arXiv preprint arXiv:1910.04867* (2019).

[542] Aston Zhang, Zachary C. Lipton, Mu Li, and Alexander J. Smola. 2020. *Dive into Deep Learning*. https://d2l.ai.

[543] Hongyi Zhang, Moustapha Cisse, Yann N Dauphin, and David Lopez-Paz. 2017. mixup: Beyond empirical risk minimization. *arXiv preprint arXiv:1710.09412* (2017).

[544] Han Zhang, Ian Goodfellow, Dimitris Metaxas, and Augustus Odena. 2018. Self-Attention Generative Adversarial Networks. arXiv:stat.ML/1805.08318

[545] Hang Zhang, Chongruo Wu, Zhongyue Zhang, Yi Zhu, Zhi-Li Zhang, Haibin Lin, Yue Sun, Tong He, Jonas Mueller, R. Manmatha, Mu Li, and Alex Smola. 2020. ResNeSt: Split-Attention Networks. *ArXiv* abs/2004.08955 (2020).

[546] Han Zhang, Zizhao Zhang, Augustus Odena, and Honglak Lee. 2019. Consistency regularization for generative adversarial networks. *arXiv preprint arXiv:1910.12027* (2019).

[547] Han Zhang, Zizhao Zhang, Augustus Odena, and Honglak Lee. 2020. Consistency Regularization for Generative Adversarial Networks. *ArXiv* abs/1910.12027 (2020).

[548] K. Zhang, Miao Sun, Tony X. Han, Xingfang Yuan, Liru Guo, and T. Liu. 2018. Residual Networks of Residual Networks: Multilevel Residual Networks. *IEEE Transactions on Circuits and Systems for Video Technology* 28 (2018), 1303–1314.

[549] Linfeng Zhang, Jiebo Song, Anni Gao, Jingwei Chen, Chenglong Bao, and Kaisheng Ma. 2019. Be Your Own Teacher: Improve the Performance of Convolutional Neural Networks via Self Distillation. arXiv:cs.LG/1905.08094

[550] Richard Zhang. 2019. Making Convolutional Networks Shift-Invariant Again. *ArXiv* abs/1904.11486 (2019).

[551] Shifeng Zhang, Cheng Chi, Yongqiang Yao, Zhen Lei, and Stan Z. Li. 2020. Bridging the Gap Between Anchor-Based and Anchor-Free Detection via Adaptive Training Sample Selection. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2020), 9756–9765.

[552] Xingcheng Zhang, Zhizhong Li, Chen Change Loy, and Dahua Lin. 2016. PolyNet: A Pursuit of Structural Diversity in Very Deep Networks. arXiv:cs.CV/1611.05725

[553] Xiaosong Zhang, Fang Wan, Chang Liu, Rongrong Ji, and Qixiang Ye. 2019. FreeAnchor: Learning to Match Anchors for Visual Object Detection. *IEEE transactions on pattern analysis and machine intelligence* PP (2019).

[554] Xiaolin Zhang, Yunchao Wei, Jiashi Feng, Yi Yang, and Thomas S. Huang. 2018. Adversarial Complementary Learning for Weakly Supervised Object Localization. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2018), 1325–1334.

[555] Xiong Zhang, Hongmin Xu, Hong Mo, Jianchao Tan, Cheng Yang, and Wenqi Ren. 2020. DCNAS: Densely Connected Neural Architecture Search for Semantic Image Segmentation. *arXiv preprint arXiv:2003.11883* (2020).

[556] Xiangyu Zhang, Xinyu Zhou, Mengxiao Lin, and Jian Sun. 2017. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. arXiv:cs.CV/1707.01083

[557] Y. Zhang and Qiang Yang. 2017. A Survey on Multi-Task Learning. *ArXiv* abs/1707.08114 (2017).

[558] Zhi Zhang, Tong He, Hang Zhang, Zhongyue Zhang, Junyuan Xie, and Mu Li. 2019. Bag of Freebies for Training Object Detection Neural Networks. *ArXiv* abs/1902.04103 (2019).

[559] Junbo Zhao, Michael Mathieu, and Yann LeCun. 2016. Energy-based Generative Adversarial Network. arXiv:cs.LG/1609.03126

[560] Shengyu Zhao, Zhijian Liu, Ji Lin, Jun-Yan Zhu, and Song Han. 2020. Differentiable Augmentation for Data-Efficient GAN Training. arXiv:cs.CV/2006.10738

[561] Shengjia Zhao, Jiaming Song, and Stefano Ermon. 2017. InfoVAE: Information Maximizing Variational Autoencoders. arXiv:cs.LG/1706.02262

[562] Stephan Zheng, Yang Song, Thomas Leung, and Ian J. Goodfellow. 2016. Improving the Robustness of Deep Neural Networks via Stability Training. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016), 4480–4488.

[563] Zhaohui Zheng, Ping Wang, Wei Liu, Jinze Li, Rongguang Ye, and Dongwei Ren. 2020. Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression. In *AAAI*.

[564] Zhedong Zheng and Yi Yang. 2021. Rectifying Pseudo Label Learning via Uncertainty Estimation for Domain Adaptive Semantic Segmentation. *International Journal of Computer Vision* (2021), 1–15.

[565] K. Zhong, Z. Song, P. Jain, Peter L. Bartlett, and I. Dhillon. 2017. Recovery Guarantees for One-hidden-layer Neural Networks. In *ICML*.

[566] Yuanyi Zhong, Jianfeng Wang, Jian Peng, and Lei Zhang. 2020. Anchor Box Optimization for Object Detection. *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)* (2020), 1275–1283.

[567] Mohan Zhou, Yalong Bai, Wei Zhang, Tiejun Zhao, and Tao Mei. 2020. Look-Into-Object: Self-Supervised Structure Modeling for Object Recognition. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2020), 11771–11780.

[568] Xingyi Zhou, Dequan Wang, and Philipp Krähenbühl. 2019. Objects as Points. *ArXiv* abs/1904.07850 (2019).

[569] Xingyi Zhou, Dequan Wang, and Philipp Krähenbühl. 2019. Objects as points. *arXiv preprint arXiv:1904.07850* (2019).

[570] Zhiming Zhou, Han Cai, Shu Rong, Yuxuan Song, Kan Ren, Weinan Zhang, Yong Yu, and Jun Wang. 2017. Activation Maximization Generative Adversarial Nets. arXiv:cs.LG/1703.02000

[571] Benjin Zhu, Jianfeng Wang, Zhengkai Jiang, Fuhang Zong, Songtao Liu, Zeming Li, and Jian Sun. 2020. AutoAssign: Differentiable Label Assignment for Dense Object Detection. *ArXiv* abs/2007.03496 (2020).

[572] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. 2017. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. arXiv:cs.CV/1703.10593

[573] Xizhou Zhu, Han Hu, Stephen Lin, and Jifeng Dai. 2019. Deformable convnets v2: More deformable, better results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 9308–9316.

[574] Xizhou Zhu, Weijie Su, Lewei Lu, Bin Li, Xiaogang Wang, and Jifeng Dai. 2021. Deformable DETR: Deformable Transformers for End-to-End Object Detection. *ArXiv* abs/2010.04159 (2021).

[575] Zhuotun Zhu, Lingxi Xie, and Alan L Yuille. 2016. Object Recognition with and without Objects. *arXiv preprint arXiv:1611.06596* (2016).

[576] Barret Zoph, Ekin Dogus Cubuk, Golnaz Ghiasi, Tsung-Yi Lin, Jonathon Shlens, and Quoc V. Le. 2020. Learning Data Augmentation Strategies for Object Detection. In *ECCV*.

[577] Barret Zoph, Golnaz Ghiasi, Tsung-Yi Lin, Yin Cui, Hanxiao Liu, Ekin Dogus Cubuk, and Quoc V. Le. 2020. Rethinking Pre-training and Self-training. *ArXiv* abs/2006.06882 (2020).

[578] Barret Zoph and Quoc V. Le. 2017. Neural Architecture Search with Reinforcement Learning. *ArXiv* abs/1611.01578 (2017).

[579] Barret Zoph, Vijay Vasudevan, Jonathon Shlens, and Quoc V. Le. 2018. Learning Transferable Architectures for Scalable Image Recognition. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2018), 8697–8710.