International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048
Volume 9, Issue 10, pp 29-36, October-2025
https://doi.org/10.47001/IRJIET/2025.910005

Neural Networks in Image Processing: A Review of Architectures, Datasets, and Performance

Mohammad Abid Al-Hashim

Department of Computer Science /College of Computer Science and Mathematics / University of Mosul, Iraq

***Abstract* - The rapid advancement of neural network-based methods has made an important transformation in image processing field. This transformation provided an unprecedented performance in a wide range of applications such as segmentation, classification enhancement, and generation. This paper provides comprehensive overview of the main neural network used in image processing, which are convolutional neural networks (CNNs), autoencoders, generative adversarial networks (GANs), and vision transformers (ViTs). The design principles behind these models have been discussed and their strengths and limitations in various image processing tasks were highlighted. Moreover, the most widely used benchmark datasets and performance metrics that facilitate objective evaluation were examined and comparison of different approaches and comparison of different approaches has been done. The trade-offs between model accuracy, computational efficiency, and scalability was also explored by analyzing recent trends. Finally, the current challenges and outline future research directions aimed at developing more efficient, interpretable, and generalizable neural network solutions for image processing have been addressed.**

Keywords: Neural Networks, Image Processing, Convolutional Neural Networks (CNNs), Image Classification, Image Segmentation, Image Enhancement, Benchmark Datasets.

I. INTRODUCTION

With the rise of neural networks, Image processing has experienced a paradigm shift, especially deep learning models, which have revolutionized tasks such as image classification, segmentation, enhancement, and generation. Early neural network models like the Neocognitron and LeNet-5 laid the groundwork for convolutional neural networks (CNNs), which became prominent because they are able to extract hierarchical spatial features from images [1][2].

The beginning of using neural networks came with the introduction of AlexNet in 2012, which has improved the performance in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) and demonstrated the effective of deep CNNs on large-scale datasets [3]. After this, CNNs have been widely adopted because of their strong inductive bias, local feature extraction capabilities, and relatively low computational requirements compared to more recent architectures [4].

However, CNNs suffer from limitations in modeling long-range dependencies and capturing global context. This was the motivation to develop new architectures such as Transformers, originally introduced for natural language processing (NLP) tasks [5]. Vision Transformers (ViTs), which adapt the self-attention mechanism to image data, have shown competitive performance on various computer vision benchmarks, often surpassing CNNs when trained on large datasets [6][7]. Alongside CNNs and ViTs, other architectures have played significant roles in advancing image processing. Autoencoders are used for unsupervised learning and image compression, while Generative Adversarial Networks (GANs) have become prominent in image synthesis, super-resolution, and style transfer tasks [8][9]. Capsule networks, although less mainstream, offer an alternative to CNNs by modeling spatial relationships and pose more explicitly [10].

Although these progress, challenges still, including high computational costs, need large annotated datasets, difficulties in model interpretability, and robustness to adversarial attacks.

This review aims to provide a comprehensive overview of the main neural network architectures used in image processing, their applications, the datasets employed for training and benchmarking, and the performance metrics used for evaluation. By synthesizing current advancements and outlining future directions, this paper serves as a resource for researchers and practitioners in the field.

II. NEURAL NETWORK ARCHITECTURES IN IMAGE PROCESSING: DETAILED OVERVIEW

In this section, the key neural network architectures that have significantly contributed to advancements in image processing have been deeply discussed, starting from classical convolutional neural networks and move towards the latest hybrid models that integrate transformer-based attention mechanisms.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 9, Issue 10, pp 29-36, October-2025

https://doi.org/10.47001/IRJIET/2025.910005

Table 1: Evolution of Neural Network Architectures in Image Processing

Architecture	Key Characteristics	Strengths	Limitations	Typical Applications	Representative Work (Ref)
CNN (Classic)	Convolution + pooling layers; local receptive fields	Effective feature extraction; efficient	Limited to local context; no global relationships	Classification, detection, segmentation	LeCun et al. (1998) [2], Krizhevsky et al. (2012) [3]
CNN (Modern Deep Variants)	Residual connections, attention modules, deep stacks	Improved expressiveness, generalization	High computation, risk of overfitting	Medical imaging, scene understanding	Khan & Iqbal (2025) [11]
Mobile CNN (e.g., MobileNet-V4)	Lightweight layers, depthwise separable convolutions	Efficient for mobile and edge computing	Less accurate than full-scale models	On-device inference, AR/VR apps	Google (2024) [12]
Autoencoder (Classic	Encoder-decoder network, often symmetric	Learns compact representations	Limited generative ability	Denoising, compression	Vincent et al. (2010) [13]
Autoencoder (Modern)	Enhanced with attention or hybrid blocks	Better reconstruction, scalability	Still outperformed by GANs/diffusion in image generation	Retrieval, anomaly detection	El-Shafai et al. (2023) [14]
GAN (Classic)	Generator-discriminator adversarial pair	Powerful image generation, creative outputs	Unstable training, mode collapse	Synthesis, inpainting	Goodfellow et al. (2014) [8]
GAN / Diffusion (Modern)	Iterative refinement or score-based generation	Realistic, high-resolution outputs	Computationally expensive	Super-resolution, generative editing	Saharia et al. (2023) [15]
Capsule Network	Encodes pose and part-whole hierarchies using dynamic routing	Handles transformations better than CNNs	Still underdeveloped, high complexity	Pose estimation, viewpoint-invariant recognition	Sabour et al. (2017) [10]
Vision Transformer (ViT)	Self-attention on image patches, no convolutions	Captures global dependencies, scalable	Requires large data and compute	Classification, segmentation	Dosovitskiy et al. (2020) [6]
Hybrid CNN–ViT	Combines local CNNs with global attention from Transformers	Hybrid CNN–ViT	Combines local CNNs with global attention from Transformers	Hybrid CNN–ViT	Combines local CNNs with global attention from Transformers

2.1 Convolutional Neural Networks (CNNs)

CNNs consider as a cornerstone of image processing since the breakthrough of LeNet-5 [2] and later AlexNet [3]. Their design depends on spatially localized receptive fields and weight sharing to efficiently capture visual patterns. their key components consist of convolutional layers for feature extraction, pooling layers for spatial down sampling, and fully connected layers for classification.

Recent CNN architectures, such as ResNet, DenseNet, and EfficientNet, include deeper networks with residual connections and optimized parameter efficiency, which enable earning of complex hierarchical features [11]. These advances improved the performance of many tasks like mage classification, object detection, and segmentation.

2.2 Mobile CNNs

Mobile CNN architectures, such as MobileNet-V4 [17], introduce depthwise separable convolutions and neural architecture search (NAS) to produce lightweight models. These models are tailored for deployment on resource-constrained devices (e.g., smartphones, drones), balancing accuracy and computational efficiency without significant loss in performance.

2.3 Autoencoders

Autoencoders are unsupervised neural networks trained to reconstruct inputs, thereby learning compressed latent representations [13]. Modern variants include transformer modules to enhance feature encoding [14]. They are used

ISSN (online): 2581-3048
Volume 9, Issue 10, pp 29-36, October-2025
https://doi.org/10.47001/IRJIET/2025.910005

widely in different applications like image denoising, compression, and anomaly detection, though their generative capabilities are limited compared to GANs.

2.4 Generative Adversarial Networks (GANs) and Diffusion Models

GANs [8] utilize adversarial training between a generator and discriminator to synthesize realistic images. Although they have challenges such as training instability, they have made a revolution in image generation and editing. Lately, powerful alternative appeared like diffusion models, which generate high-quality images through iterative refinement and score matching techniques [15].

2.5 Capsule Networks

Capsule networks aim to encode spatial hierarchies and pose relationships explicitly, addressing CNN limitations in recognizing object transformations [10]. They leverage dynamic routing algorithms between capsule units to maintain part-whole relationships. While promising, capsule networks are still computationally expensive and less widely adopted.

2.6 Vision Transformers (ViT) and Hybrid Architectures

Vision Transformers [6] apply self-attention mechanisms on image patches, allowing models to capture global dependencies beyond the local receptive fields of CNNs. However, ViTs generally require large-scale datasets and considerable compute.

To combine local feature extraction and global context modeling, hybrid CNN–ViT architectures have been proposed [16]. These models utilize convolutional layers to extract local features followed by transformer blocks for global reasoning, achieving strong performance across a range of vision tasks.

III. DATASETS AND BENCHMARKS IN IMAGE PROCESSING

Datasets play a crucial role in the development and evaluation of neural network architectures for image processing. High-quality, diverse datasets enable models to generalize better and provide standardized benchmarks to compare different approaches objectively.

3.1 Popular Datasets for Image Classification

- **ImageNet:** One of the largest and most widely used datasets for image classification, containing over 14 million labeled images across 21,000+ categories. The ImageNet Large Scale Visual Recognition Challenge (ILSVRC) has driven significant progress in CNN architectures [18].
- **CIFAR-10 and CIFAR-100:** Small-scale datasets with 60,000 images in 10 and 100 classes respectively. They are popular for benchmarking lightweight and experimental models due to their manageable size and complexity [19].
- **Tiny ImageNet:** A scaled-down subset of ImageNet with 200 classes and 100,000 images, used to test models in constrained environments [20].

3.2 Datasets for Object Detection and Segmentation

- **COCO (Common Objects in Context):** Contains over 330,000 images with over 2.5 million labeled instances spanning 80 object categories. COCO challenges focus on object detection, segmentation, and captioning tasks [21].
- **Pascal VOC:** Offers annotated images for object detection and segmentation with 20 categories. It serves as a foundational benchmark for object recognition algorithms [22].
- **Cityscapes:** Focused on urban scene understanding, it provides high-quality pixel-level annotations for semantic segmentation in street scenes [23].

3.3 Datasets for Image Generation and Reconstruction

- **CelebA:** A large-scale face attributes dataset with over 200,000 celebrity images, used extensively in GAN and autoencoder research for face synthesis and editing [24].
- **LSUN:** Large-scale scene understanding dataset with millions of images across various scene categories, commonly used for generative modeling [25].

3.4 Emerging Datasets

- **ImageNet-V2 and ImageNet-R:** Variants of ImageNet designed to evaluate robustness and domain shifts [26].
- **Open Images Dataset V7:** A massive dataset with ~9 million images annotated with image-level labels, object bounding boxes, and segmentation masks, supporting various vision tasks [27].

Table 2: Commonly Used Datasets in Image Processing

Dataset	Task(s)	Size	Key Features	Reference
ImageNet	Classification	14M+ images, 21k classes	Large-scale, diverse classes	[18]
CIFAR-10/100	Classification	60k images	Small-scale, simple classes	[3]
COCO	Detection, segmentation	330k images	Multiple objects, instance segmentation	[21]

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 9, Issue 10, pp 29-36, October-2025

https://doi.org/10.47001/IRJIET/2025.910005

Pascal VOC	Detection, segmentation	20k images	Well-annotated, classical benchmark	[22]
CelebA	Image generation	200k images	Facial attributes, face synthesis	[24]
LSUN	Image generation	Millions of images	Large-scale scene categories	[25]

IV. PERFORMANCE METRICS AND EVALUATION METHODS

Evaluating neural networks in image processing requires robust and relevant metrics tailored to specific tasks such as classification, detection, segmentation, and generation. This section summarizes the most commonly used performance metrics and evaluation protocols.

4.1 Metrics for Image Classification

1. Accuracy:

The proportion of correctly classified samples out of the total and it measures the overall correctness of a classifier. It is simple and widely used but can be misleading in imbalanced datasets [18].Equation **(**1) represents the formula used to determine accuracy:

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (1)$$

Where: TP represents True Positives, TN is True Negatives, FP is False Positives, FN is False Negatives.

2. Precision, Recall, and F1-Score:

Precision measures the proportion of true positive predictions among all positive predictions. So, it represents the proportion of positive identifications that were actually correct. Equation (2) used to determine accuracy:

$$precision = \frac{TP}{TP+FP} \qquad (2)$$

Recall (or sensitivity) measures the proportion of true positives detected among all actual positives. It represents the proportion of actual positives that were identified correctly and it can be calculated using equation (3).

$$recall = \frac{TP}{TP+FN} \qquad (3)$$

F1-Score is the harmonic mean of precision and recall, balancing both metrics [28]. The mathematic formula used to determine F1-Score shown in equation (4):

$$F1 = 2 \times \frac{precision \times recall}{precision + recall} \qquad (4)$$

3. Top-k Accuracy

Top-k Accuracy: Measures whether the true class is among the model's top k predicted probabilities. Commonly used in ImageNet evaluations (e.g., top-1 and top-5 accuracy) [18].

4.2 Metrics for Object Detection

Mean Average Precision (mAP): The average precision across all classes and multiple Intersection over Union (IoU) thresholds. It summarizes precision-recall curves and is the primary metric for object detection challenges like COCO and Pascal VOC [22,29].often computed via area under the precision-recall curve.

Intersection over Union (IoU): Measures overlap between predicted bounding boxes and ground truth. IoU thresholds (e.g., 0.5 or 0.75) define true positives in detection tasks [22]. It can be determined using formula in equation (5).

$$IoU = \frac{B_p \cap B_p}{B_p \cup B_{gt}} \qquad (5)$$

4.3 Metrics for Semantic and Instance Segmentation

- **Mean Intersection over Union (mIoU):** Extends IoU for pixel-wise classification, averaging over all classes. A higher mIoU indicates better segmentation quality [29].
- **Pixel Accuracy:** The ratio of correctly classified pixels to the total pixels [29].

4.4 Metrics for Image Generation and Reconstruction

- **Peak Signal-to-Noise Ratio (PSNR):** Quantifies reconstruction quality by comparing the similarity between generated and reference images. Higher PSNR indicates better quality [28]. Equation (6) represents the formula used to determine it:

$$PSNR = 10 \times \log_{10}\left(\frac{MAX_I^2}{MSE}\right) \qquad (6)$$

Where MAXI is the maximum possible pixel value (e.g., 255 for 8-bit images), MSE is mean squared error between the generated and reference images.

- **Structural Similarity Index Measure (SSIM):** Assesses perceptual similarity based on luminance, contrast, and structure [9]. SSIM values range from -1 to 1, where 1 means perfect similarity. Equation (7) shows its mathematical determination.

ISSN (online): 2581-3048
Volume 9, Issue 10, pp 29-36, October-2025
https://doi.org/10.47001/IRJIET/2025.910005

$$SSIM\ (x,y) = \frac{(2\mu_x\,\mu_y+C_1)(2\sigma_{xy}+C_2)}{(\mu_x^2+\mu_y^2+C_1)(\sigma_x^2+\sigma_y^2+C_2} \qquad (7)$$

Where μ_x, μ_y represent mean of x and y, σ_x^2, σ_y^2 are the variance of x and y ,σ_{xy} is the covariance of x and y and C_1 , C_2 are the stabilization constants.

- **Fréchet Inception Distance (FID):** Measures similarity between distributions of generated images and real images using feature embeddings from pretrained networks. Lower FID indicates better quality and diversity [30].
- **Inception Score (IS):** Evaluates the diversity and quality of generated images using a pretrained classifier [31].

4.5 Evaluation Protocols

- **Cross-validation:** Commonly used for smaller datasets to reduce variance in performance estimates [28].
- **Train/Validation/Test Splits:** Ensures models are evaluated on unseen data, preventing overfitting [8].
- **Benchmark Challenges:** Standardized competitions such as ILSVRC, COCO challenges, and ImageNet Robustness benchmarks provide consistent evaluation frameworks [18,21].

V. APPLICATIONS OF NEURAL NETWORKS IN IMAGE PROCESSING

Neural networks have revolutionized a wide range of image processing tasks by automating feature extraction and achieving superior performance over traditional methods. This section provides an overview of the most impactful applications, categorized by task.

5.1 Image Classification

Image classification considers as one of the earliest and most well-established applications of neural networks. In this part convolutional Neural Networks (CNNs) are very effective because of their ability to learn spatial hierarchies of features. The accuracy of image classification has improved significantly on benchmark datasets such as ImageNet [18] from beginning models like AlexNet [32] to recent architectures like EfficientNet [33] and Vision Transformers [6]

5.2 Object Detection

Object detection involves identifying and localizing multiple objects in an image. Models like Faster R-CNN [34], YOLO (You Only Look Once) [35], and more recently YOLOv8 [36] offer real-time performance with high accuracy. These models are widely used in surveillance, autonomous vehicles, and industrial automation.

5.3 Image Segmentation

Segmentation is the process of partitioning an image into meaningful regions:

- **Semantic segmentation** assigns a class label to each pixel. Popular models include U-Net [37], DeepLabV3+ [38], and SegFormer [39].
- **Instance segmentation** combines object detection and semantic segmentation, labeling each instance separately. Models like Mask R-CNN [40] are commonly used here.

Applications include medical imaging, autonomous driving, and agriculture.

5.4 Image Super-Resolution

Super-resolution enhances the resolution of an image using deep learning. CNNs and GAN-based architectures like SRGAN [41] and ESRGAN [9] learn to reconstruct fine image details from low-resolution inputs. More recently, diffusion-based models have been applied to this task [15].

5.5 Image Generation and Editing

Generative Adversarial Networks (GANs) and diffusion models have enabled photorealistic image synthesis, inpainting, style transfer, and face editing. Notable models include StyleGAN [42] for high-quality facial image generation and DALL•E for text-to-image synthesis [43].

5.6 Anomaly Detection

Autoencoders and GANs are widely used for visual anomaly detection in industrial inspection, medical imaging, and video surveillance. They learn a distribution of normal data and identify deviations from this distribution during inference [44].

VI. CHALLENGES AND FUTURE DIRECTIONS IN NEURAL NETWORK–BASED IMAGE PROCESSING

Despite remarkable progress, applying neural networks to image processing still faces several open challenges. Understanding these issues is key to identifying research gaps and guiding the development of next-generation models.

6.1 Current Challenges

6.1.1 Data Limitations and Annotation Cost

Deep learning models are data-hungry. High-quality, labeled datasets are often expensive and time-consuming to curate especially in domains like medical imaging or remote sensing where expert annotation is required [45]. Furthermore,

ISSN (online): 2581-3048
Volume 9, Issue 10, pp 29-36, October-2025
https://doi.org/10.47001/IRJIET/2025.910005

datasets may not cover sufficient diversity, leading to poor generalization.

6.1.2 Generalization and Robustness

Models trained on specific datasets often fail to generalize well to real-world, unseen scenarios a phenomenon known as dataset bias. Even small perturbations (e.g., adversarial noise, lighting changes) can significantly affect model performance [46].

6.1.3 Computational Cost and Energy Consumption

State-of-the-art models (e.g., Vision Transformers, large CNNs) require significant computational power for training and inference, limiting their deployment on edge devices or in real-time applications. This also raises environmental concerns due to energy usage [47].

6.1.4 Interpretability and Explainability

Neural networks are often viewed as "black boxes." Lack of interpretability makes it difficult to trust model predictions in critical applications such as healthcare, law enforcement, or autonomous vehicles [48].

6.1.5 Model Bias and Ethical Concerns

Biased training data can lead to discriminatory outcomes, especially in facial recognition or surveillance applications. Ethical concerns also extend to the use of generative models for misinformation and privacy violations [5].

VII. CONCLUSION

Neural networks have profoundly transformed the field of image processing, enabling unprecedented advances in tasks such as classification, detection, segmentation, super-resolution, and image synthesis. From early convolutional architectures to contemporary transformer-based and diffusion models, the evolution of network designs has pushed the boundaries of performance and application scope.

In this review, we have surveyed the key neural network architectures used in image processing, benchmark datasets that have shaped research progress, and standard performance metrics used for evaluation. We also highlighted real-world applications across domains including healthcare, autonomous driving, and industrial inspection, alongside current limitations and emerging research directions.

Despite notable successes, several challenges remain — particularly in terms of data efficiency, model generalization, interpretability, and ethical concerns. Addressing these issues will be essential for deploying neural networks responsibly and effectively in high-stakes, real-world environments.

REFERENCES

[1] Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36(4), 193–202.

[2] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2324.

[3] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *In Advances in Neural Information Processing Systems (NeurIPS)*, 25, 1097–1105.

[4] Rawat, W., & Wang, Z. (2017). Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. Neural Computation, 29(9), 2352–2449.

[5] Vaswani, A., et al. (2017). Attention is All You Need. *In Advances in Neural Information Processing Systems (NeurIPS)*, 30.

[6] Dosovitskiy, A., et al. (2020). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv preprint* arXiv:2010.11929.

[7] Khan, S., Naseer, M., Hayat, M., Zamir, S. W., Khan, F. S., & Shah, M. (2022). Transformers in Vision: A Survey. *ACM Computing Surveys*, 54(10), 1–41.

[8] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. *In Advances in neural information processing systems* (pp. 2672-2680).

[9] Wang, X., Yu, K., Dong, C., Loy, C. C. (2018). Recovering realistic texture in image super-resolution by deep spatial feature transform. *In IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 606–615.

[10] Sabour, S., Frosst, N., & Hinton, G. E. (2017). Dynamic Routing Between Capsules. *In Advances in Neural Information Processing Systems (NeurIPS)*, 30.

[11] Khan, S., & Iqbal, R. (2025). A comprehensive survey on architectural advances in deep CNNs: Challenges, applications, and emerging research directions. *arXivpreprint arXiv*:2503.16546. https://arxiv.org/abs/2503.16546

[12] Google (2024). MobileNet V4. Wikipedia

[13] Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., & Manzagol, P. A. (2010). Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of Machine Learning Research*, 11, 3371-3408.

ISSN (online): 2581-3048
Volume 9, Issue 10, pp 29-36, October-2025
https://doi.org/10.47001/IRJIET/2025.910005

[14] El-Shafai, W. E., et al. (2023). Image retrieval using convolutional autoencoder, InfoGAN, and vision transformer unsupervised models. *ResearchGate.* https://www.researchgate.net/publication/368234541

[15] Saharia, C., et al. (2023). Image Super-Resolution via Iterative Refinement. *IEEE TPAMI.*

[16] Yunusa, H., et al. (2024). Hybrid CNN–ViT Architectures for Computer Vision. *arXiv:*2402.02941.

[17] Howard, A., et al. (2024). MobileNetV4: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint* arXiv:2403.XXXX. [Note: Use actual link or publication info if available]

[18] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–255. https://doi.org/10.1109/CVPR.2009.5206848

[19] Krizhevsky, A. (2009). Learning multiple layers of features from tiny images. *Technical Report, University of Toronto*. http://www.cs.toronto.edu/~kriz/cifar.html

[20] Tiny ImageNet Challenge. (2015). Stanford CS231n: Convolutional Neural Networks for Visual Recognition. http://tiny-imagenet.herokuapp.com/

[21] Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., ... & Zitnick, C. L. (2014). Microsoft COCO: Common objects in context. *European Conference on Computer Vision*, 740–755. https://doi.org/10.1007/978-3-319-10602-1_48

[22] Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., & Zisserman, A. (2010). The Pascal Visual Object Classes (VOC) challenge. *International Journal of Computer Vision*, 88(2), 303–338. https://doi.org/10.1007/s11263-009-0275-4

[23] Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., ... & Schiele, B. (2016). The Cityscapes dataset for semantic urban scene understanding. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3213–3223. https://doi.org/10.1109/CVPR.2016.350

[24] Liu, Z., Luo, P., Wang, X., & Tang, X. (2015). Deep learning face attributes in the wild. *Proceedings of the IEEE International Conference on Computer Vision*, 3730–3738. https://doi.org/10.1109/ICCV.2015.425

[25] Yu, F., Zhang, Y., Song, S., Seff, A., & Xiao, J. (2015). LSUN: Construction of a large-scale image dataset using deep learning with humans in the loop. *arXiv preprint arXiv*:1506.03365. https://arxiv.org/abs/1506.03365

[26] Recht, B., Roelofs, R., Schmidt, L., & Shankar, V. (2019). Do ImageNet classifiers generalize to ImageNet? *arXiv preprint arXiv*:1902.10811. https://arxiv.org/abs/1902.10811

[27] Open Images Dataset V7. (2023). Google Research. https://storage.googleapis.com/openimages/web/index.html

[28] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," *in Proc. 14th Int. Joint Conf. Artif. Intell. (IJCAI), Montreal, Canada*, 1995, pp. 1137–1143.

[29] T.-Y. Lin et al., "Microsoft COCO: Common objects in context," *in Proc. Eur. Conf. Comput. Vis. (ECCV), Zurich, Switzerland*, 2014, pp. 740–755.

[30] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs trained by a two time-scale update rule converge to a local Nash equilibrium," *in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), Long Beach, CA, USA*, 2017, pp. 6626–6637.

[31] C. Salimans et al., "Improved techniques for training GANs," *in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), Barcelona, Spain*, 2016, pp. 2234–2242.

[32] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25, 1097–1105.

[33] Tan, M., & Le, Q. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. *Proceedings of the 36th International Conference on Machine Learning*, 6105–6114.

[34] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems*, 28.

[35] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 779–788.

[36] Jocher, G., Chaurasia, A., & Qiu, J. (2023). YOLOv8: A cutting-edge object detection and segmentation model. *Ultralytics Technical Report.* https://github.com/ultralytics/ultralytics

[37] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 234–241.

[38] Chen, L. C., Zhu, Y., Papandreou, G., et al. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. *European Conference on Computer Vision (ECCV)*, 801–818.

[39] Xie, E., Wang, W., Yu, Z., et al. (2021). SegFormer: Simple and efficient design for semantic segmentation with transformers. *Advances in Neural Information Processing Systems*, 34.

ISSN (online): 2581-3048
Volume 9, Issue 10, pp 29-36, October-2025
https://doi.org/10.47001/IRJIET/2025.910005

[40] He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2961–2969

[41] Ledig, C., Theis, L., Huszár, F., et al. (2017). Photo-realistic single image super-resolution using a generative adversarial network. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 4681–4690.

[42] Karras, T., Laine, S., & Aila, T. (2019). A style-based generator architecture for generative adversarial networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 4401–4410.

[43] Ramesh, A., Pavlov, M., Goh, G., et al. (2021). Zero-shot text-to-image generation. *International Conference on Machine Learning (ICML).*

[44] Schlegl, T., Seeböck, P., Waldstein, S. M., et al. (2017). Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. *Information Processing in Medical Imaging (IPMI)*, 146–157.

[45] Tajbakhsh, N., Jeyaseelan, L., Li, Q., Chiang, J. N., Wu, Z., & Ding, X. (2020). Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. *Medical Image Analysis*, 63, 101693. https://doi.org/10.1016/j.media.2020.101693

[46] Hendrycks, D., & Dietterich, T. (2019). Benchmarking neural network robustness to common corruptions and perturbations. *International Conference on Learning Representations (ICLR).*

[47] Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 3645–3650.

[48] Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint* arXiv:1702.08608. https://arxiv.org/abs/1702.08608

[49] Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research*, 81, 1–15.

AUTHOR'S BIOGRAPHY

Mohammad Abid Al-Hashim received the B.Sc. degree incomputer science from Collage of Computer Science and Mathematics / University of Mosul, Iraq. The M.Sc. degree in computer science from Collage of Computer Science and mathematics / University of Mosul, Iraq. He is an assistant lecturer at University of Mosul, Iraq. He can be contacted at email: maqassim@uomosul.edu.iq

Citation of this Article:

Mohammad Abid Al-Hashim. (2025). Neural Networks in Image Processing: A Review of Architectures, Datasets, and Performance. *International Research Journal of Innovations in Engineering and Technology - IRJIET*, 9(10), 29-36. Article DOI https://doi.org/10.47001/IRJIET/2025.910005

****

