Deep Learning: Semester Assignment 2025

Nikolaos Theokritos Tsopanidis aivc24022 22/06/2025

Περιεχόμενα

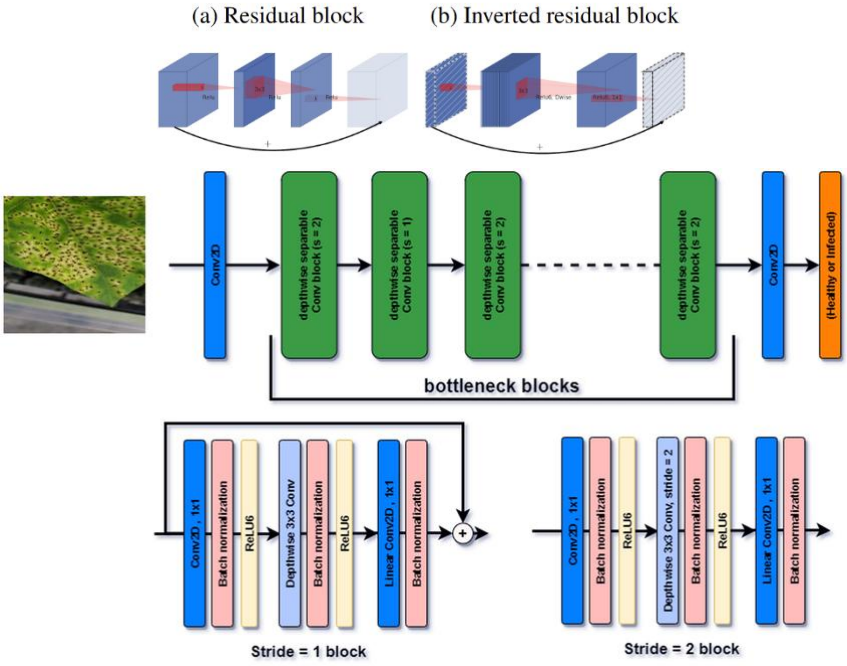
Part 1

For the first part of the work, the MobileNetV2 and EfficientNetB0 models were used in the categorization of the images of the CIFAR-100 dataset. The models were chosen mainly for the small number of parameters they have compared to other deep convolutional models of computer vision, but also for the effectiveness they have shown due to their innovative architecture and mechanisms. Below is an analysis of these models and their performance in the CIFAR-100 dataset is discussed.

MobileNetV2

MobileNetV2 is a highly efficient architecture specifically designed for mobile and embedded vision applications. It was developed by Google researchers as an improvement over the original MobileNet model, with the aim of achieving competitive accuracy while reducing computational costs and model size [1].

The architecture of MobileNetV2 is based on a series of convolutional layers, with the Inverted Residual Block as a basic element. Unlike traditional residual blocks that extend the feature map and then constrain it, MobileNetV2's inverted residual block does the opposite, connecting thin bottleneck layers as the entrance and exit of the block, with an intermediate expansion layer that uses light deep convolutions.



The architecture of MobileNetV2 with a sample input image and residual congestion levels. [2]

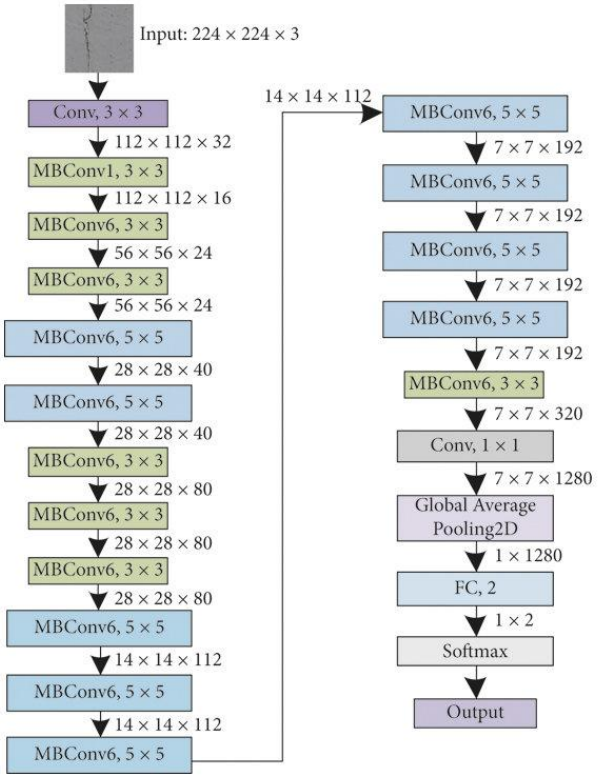
1. The expansion layer: A 1x1 convolution that increases the number of channels (also known as the expansion factor). This level is usually followed by a ReLU6 trigger function to introduce nonlinearity.
2. Depthwise Seperable Convolution: A lightweight in-depth convolution that filters features in the extended representation. This feature applies only one filter per input channel, significantly reducing the calculation compared to standard convolutions.

3. Projection Level: A 1x1 convolution that projects the high-dimensional features back to a lower-dimensional output, forming the congestion level.

The overall architecture of MobileNetV2 includes an initial fully convolutional layer with 32 filters, followed by 19 residual congestion levels.

EfficientNetB0

EfficientNetB0 is the core model of the EfficientNet family, which are designed to achieve high performance with significantly fewer computing resources and parameters compared to previous architectures. It was introduced by Mingxing Tan and Quoc V. LE in 2019 [3] and its architecture is mainly based on mobile inverted congestion convolutions (MBConv), where it integrates compression and excitation optimization (SE | squeeze-and-excitation).



EfficientnetB0 Model Architecture [4]

The EfficientNet-B0 network consists of:

1. The initial convolution layer followed by batch normalization and a ReLU6 activation function,
2. A series of MBConv blocks which are the basic building blocks and feature in-depth separable convolutions to reduce computational costs, compression and stimulation layers to improve the representation of features through the recalibration of responses per channel,
3. and a final convolutional block, followed by a Global Average Pooling layer and a dense output layer with a Softmax trigger function for sorting.

EfficientNet models, including EfficientNetB0, demonstrate strong performance across various datasets, including CIFAR-100. The initial study reported that EfficientNetB0 achieved a classification accuracy of

88.1%, while reducing parameters. Other studies and implementations for CIFAR-100 using EfficientNetB0 can present a variety of results based on training configurations, such as number of seasons, learning rate, data augmentation, etc.

Model training

For training in the CIFAR-100 dataset, the pre-trained models were initialized using TensorFlow as the core framework.

70% Training Set

	Accuracy (%)	**Precision (%)**	**Recall (%)**	**F1-Score (%)**	**Training Time**
MobileNetV2	64.37	80.52	52.44	64.14	10m 1s
EfficientNetB0	73.29	86.69	63.24	73.22	12m 22s

60% Training Set

	Accuracy (%)	**Precision (%)**	**Recall (%)**	**F1-Score (%)**	**Training Time**
MobileNetV2	64.05	79.47	52.85	63.88	8m 43s
EfficientNetB0	71.39	86.17	59.95	71.31	10m 37s

50% Training Set

	Accuracy (%)	**Precision (%)**	**Recall (%)**	**F1-Score (%)**	**Training Time**
MobileNetV2	62.91	79.5	50.84	62.71	7m 29s
EfficientNetB0	70.69	85.96	58.77	70.64	8m 54s

Analyzing the results of the categorization, we can conclude that EfficientNetB0 consistently demonstrates superior performance in all measurements, recording longer training times due to increased computing costs. MobileNetV2 showed poorer results but showed greater training speed, highlighting its preference in scenarios where we have limited computing resources. As expected, for both models, performance decreases as the size of the training set shrinks. The decrease in accuracy is steady, proving the fundamental principle of machine learning that more data leads to better and more generalized models.

Part 2

PanNuke Dataset

The PanNuke dataset is an extensive, available resource for training models for the segmentation and classification of cell nuclei in hematoxylin and eosine-stained histopathology images. It was developed by Gamper et al. (2020) and responds to the need for automated core analysis in cancer research. [5]

It contains 7,901 patches of images with a resolution of 256x256 pixels, in PNG format, from 19 different types of tissues. Each image segment comes with a corresponding mask that classifies pixels into one of five types of nuclei (neoplastic, inflammatory, connective tissue, dead, and epithelial) or on a background. The initial dataset is divided into 60% ratios for training, 20% for validation, and 20% for testing.

PanNuke addresses the challenge of automatically delimiting and categorizing individual nuclei in images of cancer tissue. This is important for understanding the tumor, aiding in diagnosis, and determining treatment, as manual analysis is time-consuming and prone to variability. Its pan-cancer field allows for the creation of models that generalize to different types of cancer. It is used as a reference point for deep learning models in computational pathology, supporting tasks such as identifying specific types of nuclei in diagnostic workflows and studying cellular interactions.

Models such as Hover-Net, and LKCell-L have been evaluated in PanNuke, with performance measured using Pan-Optical Quality (PQ) as a metric for the combined quality of segmentation and classification. For example, Hover-Net achieved binary PQ 0.6596 and multi-class PQ 0.4629, while LKCell-L displayed binary PQ 0.6847 and multi-class PQ 0.5080. [6] Additional models that have been evaluated with the Dice Coefficient metric are SONNET and HDA-Net which scored 0.824 and 0.8203 respectively. [7, 8]

PanNuke stands out due to its large scale (over 200,000 annotated cores) and its diversity across 19 tissue types, making it a comprehensive dataset like CoNSeP or MoNuSeg.

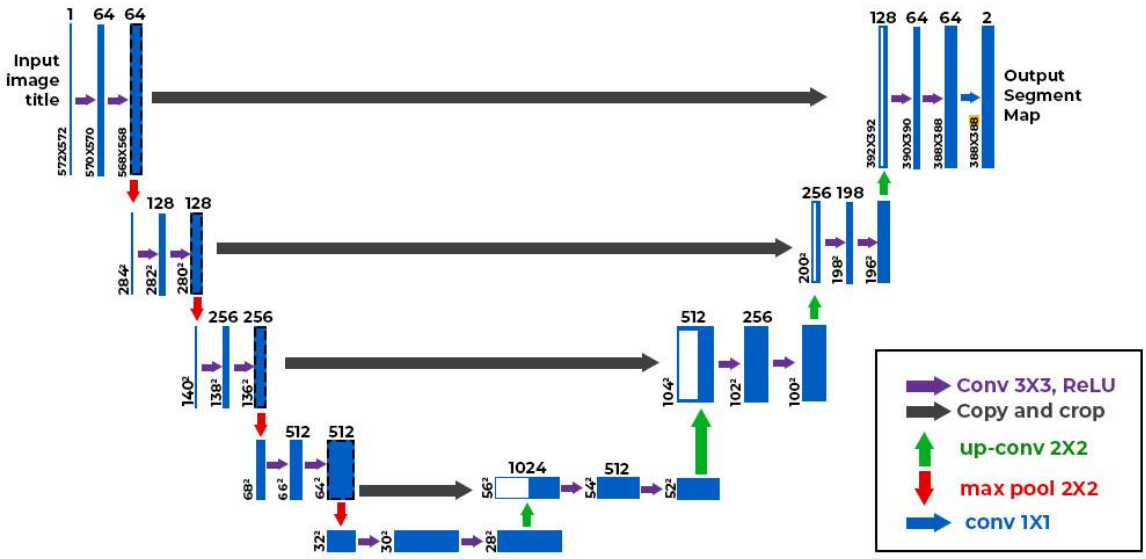
My choice to use the PanNuke dataset for the work has direct and deep roots in my background as a biomedical engineer, particularly in my interest and experience in medical imaging and computational analysis of biological data. Our role often involves bridging the gap between biological and medical problems and technological solutions. In this context, the precise segmentation and classification of cell nuclei are important tasks for quantitative histopathology.

UNet & DeepLabV3+ Analysis

Part of the work is the analysis and comparison of two prominent deep learning architectures in computer vision, which are used extensively in image segmentation work, especially in the field of biology and medicine [9, 10], UNet and DeepLabV3+. The segmentation of medical images has proven to be a very important issue for the development of treatments, diagnostic methods and research, a process that is achieved with great efficiency using the aforementioned computer vision models. Below, the model architectures will be analyzed and evaluated based on performance metrics in the PanNuke dataset segment.

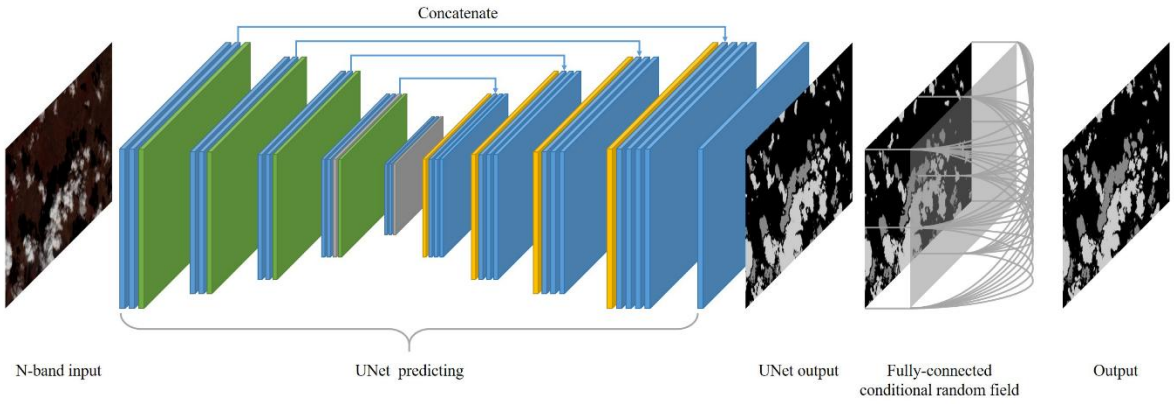
U-Net

U-Net is a convolutional neural network designed in 2015 for biomedical image fragmentation [9], and characterized by a symmetrical encoder-decoder structure. The main innovation of the architecture is the "U" shaped design including Shortcut Connections.



These connections directly transfer high-resolution feature "maps" from the encoder to the decoder, retaining detailed spatial information that is lost during sampling. This leads to accurate object detection and sharper fragmentation boundaries, which is important for delimiting structures such as cell membranes or tumor boundaries [11].

The two main parts of the architecture are the codec and the decoder. The encoder extracts hierarchical characteristics and contextual information through convolutional layers and "Max-Polling" sampling. The decoder reconstructs the partition "map" through oversampling and over-convolution. The oversampled features are linked to high-resolution features by the encoder through the bypass connections, and then the subsequent convolutional layers are improved.



U-Net is highly effective, performing well even on small datasets, which is common in medical imaging due to privacy, difficult annotations, and the need for high expertise. This efficiency, combined with data augmentation, allows it to learn valid characteristics from limited data. Variants such as nnU-Net and TransUNet further enhance its adaptability [12, 13].

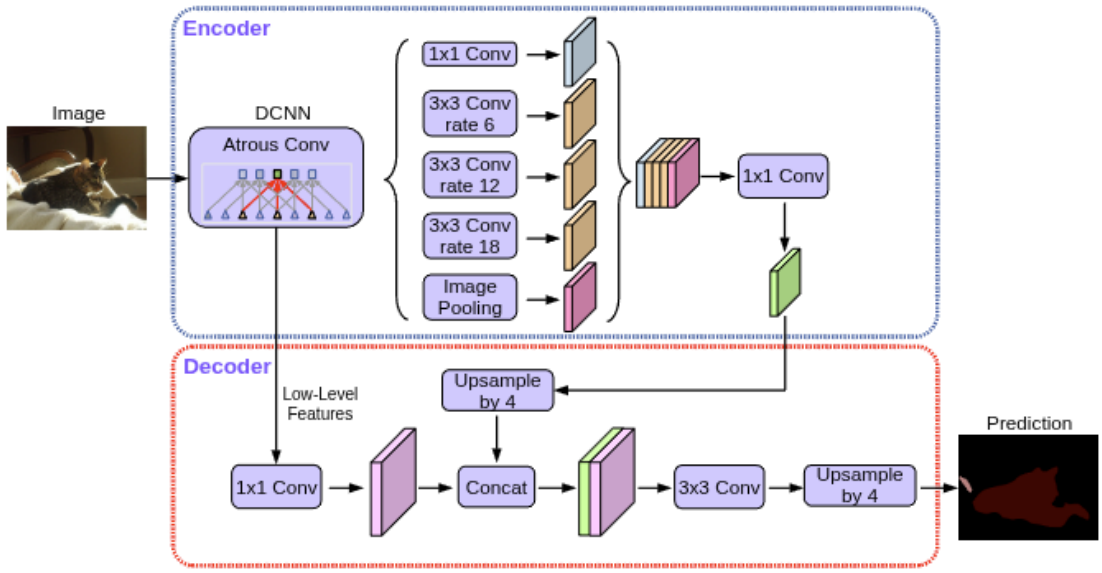
U-Net has played an important role in various biomedical applications such as cell segmentation, where it was used for accurate classification of platelets per pixel [14]. To segment retinal vessels, aiding in early detection of diseases in ophthalmology [15], to segment cancerous tumors using variants of the architecture, such as 3D-UNet and RA-Unet [16, 17] and to segment melanoma using UNet with DenseNet-77 as an

encoder [18]. Relevant to the present work is the study by C. Tommasino et al. (2023), where the HoVer-UNet model was designed using the U-Net architecture with the aMix Vision Transformer as the backbone of the overall model. The model was trained on the PanNuke and Concep datasets and was used in multi-category core segmentation, while achieving decent results, outperforming widely used models such as DIST and Mask-RCNN in performance [19].

DeepLabV3+

DeepLabV3+ improves semantic image segmentation by combining spatial pyramid grouping and coder-decoder structures. Its main innovation is the integration of Atrous Spatial Pyramid Pooling (ASPP) with an efficient decoder model. [20]

The architectural features of DeepLabV3+ typically include a pre-trained backbone as an encoder (e.g., ResNet, Xception) to extract high-level features, the ASPP that processes the encoder output, and a simple but efficient decoder.



ASPP allows you to capture frame information at multiple scales, which is especially useful for segmenting objects that vary greatly in size within an image, as is often the case in biomedical images. At the same time, it applies many atroic convolutions that allow us to control the analysis in which the responses of the features within the encoder are calculated. [21]

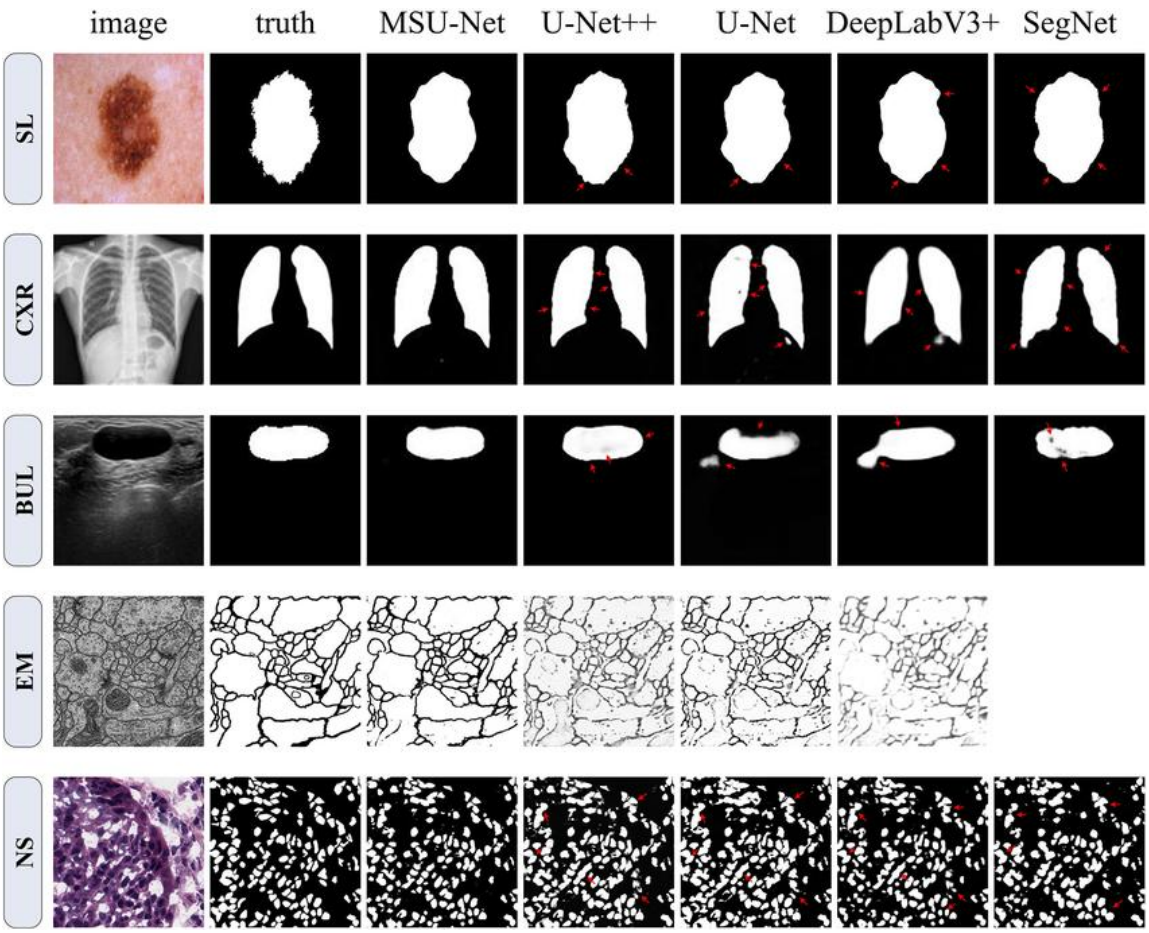
DeepLabV3+ has shown strong performance in biomedical imaging through various implementations and applications that have been made, such as cell segmentation [22], cancerous tumors – achieving high performance in datasets such as BraTS [23], melanoma segmentation [24] etc.

Comparison in biomedical image segmentation

When we compare U-Net and DeepLabV3+ for biomedical image segmentation, various points of analysis emerge. U-Net often excels at accurate demarcation due to the direct detours that retain detailed spatial

information. This makes it particularly powerful for segmenting small, complex objects where precise boundaries are important. DeepLabV3+, with ASPP, is excellent at capturing multi-scale environment information, which can be beneficial for segmenting objects with a wide variety of sizes or for tasks where the global environment is more important. [25]

In terms of the computational complexity of the models, DeepLabV3+ can sometimes be more expensive due to its advanced core network and ASPP, although atroic convolutions help alleviate this. The U-Net, while effective, can be considered "shaltier" in some contexts, allowing for faster training on limited material. In conclusion, both U-Net and DeepLabV3+ are powerful tools for image partitioning, each with its own unique advantages.



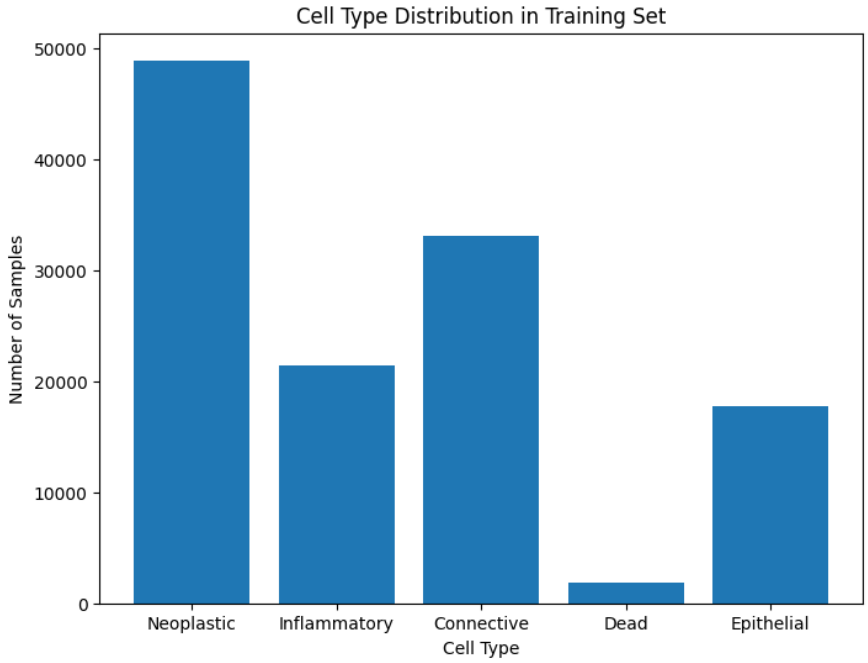
Quality comparison between SegNet, Deeplabv3, U-Net++, U-Net, and MSU-Net. The application of various models in biomedical image segmentation is presented. The red arrows show where the models have made the wrong segmentation. [26]

Semantic Partition of the PanNuke Set

This module describes the experimental process and the results of the semantic partitioning performed with three architectures, a simple U-Net, a U-Net with ResNet50, and a DeepLabV3+ with ResNet50. ResNet50 was chosen as the encoder of the models to extract higher quality high-dimensional features. The models were trained and evaluated to compare their effectiveness in identifying different types of cell nuclei.

Dataset Pre-Processing

The PanNuke set was acquired from the Hugging Face website and was structured into three sub-groups (fold1, fold2, fold3). Each of these sub-groups contained the images and corresponding masks of the cell nuclei. For the training of the models, the fold1 and fold2 sub-groups were combined to form a total of 5,179 samples for the training set. The fold3 was divided evenly, creating the validation and control subsets, each comprising 1,361 samples. To address the known dataset imbalance between classes, particularly in the dead cell category, weights were included in training.



Category	Background	Neoplastic	Inflammatory	Connective	Dead	Epithelial
Number of samples	N/A	48.932	21.451	33.144	1.851	17.712
Weight	1.0000	1.5031	2.1476	1.7428	14.2998	2.3899

Education

For the training process, a graphics card with powerful computing capabilities was used, NVIDIA GeForce RTX 4070 with 12GB memory, allowing for quick training of resource-intensive models and exploitation of batches of data. Various batch size values were tested and it was decided that each batch should include 4 samples as the optimal choice due to the training speed and memory usage rate of the graphics card. normalized using ImageNet's average RGB value [0.485, 0.456, 0.406] and standard deviation [0.229, 0.224, 0.225] and small data increments were applied by slightly changing the sharpness, brightness, contrast, saturation, and hue of the images. The Adam algorithm was used as an optimizer and CrossEntropy as an error function as common practices in computer vision and multi-category structure segmentation tasks. An early termination mechanism was implemented when the loss in the validation set did not improve for the 12th era. The original goal was 100 seasons. The performance of the models was evaluated with the Mean Intersection over Union (mIoU), Dice Coefficient, Pixel Accuracy, Precision, and Recall metrics.

Results

Train

Model	Training time	Epochs	Loss	Pixel Accuracy	Precision	Recall	mIoU	Dice
Vanilla U-Net	86m12s	34	0.360	0.905	0.639	0.644	0.495	0.610
ResNet50 + U-Net	38m44s	28	0.372	0.903	0.640	0.643	0.493	0.606
ResNet50 + DeepLabV3+	31m27s	24	0.371	0.902	0.641	0.637	0.491	0.604

Validation

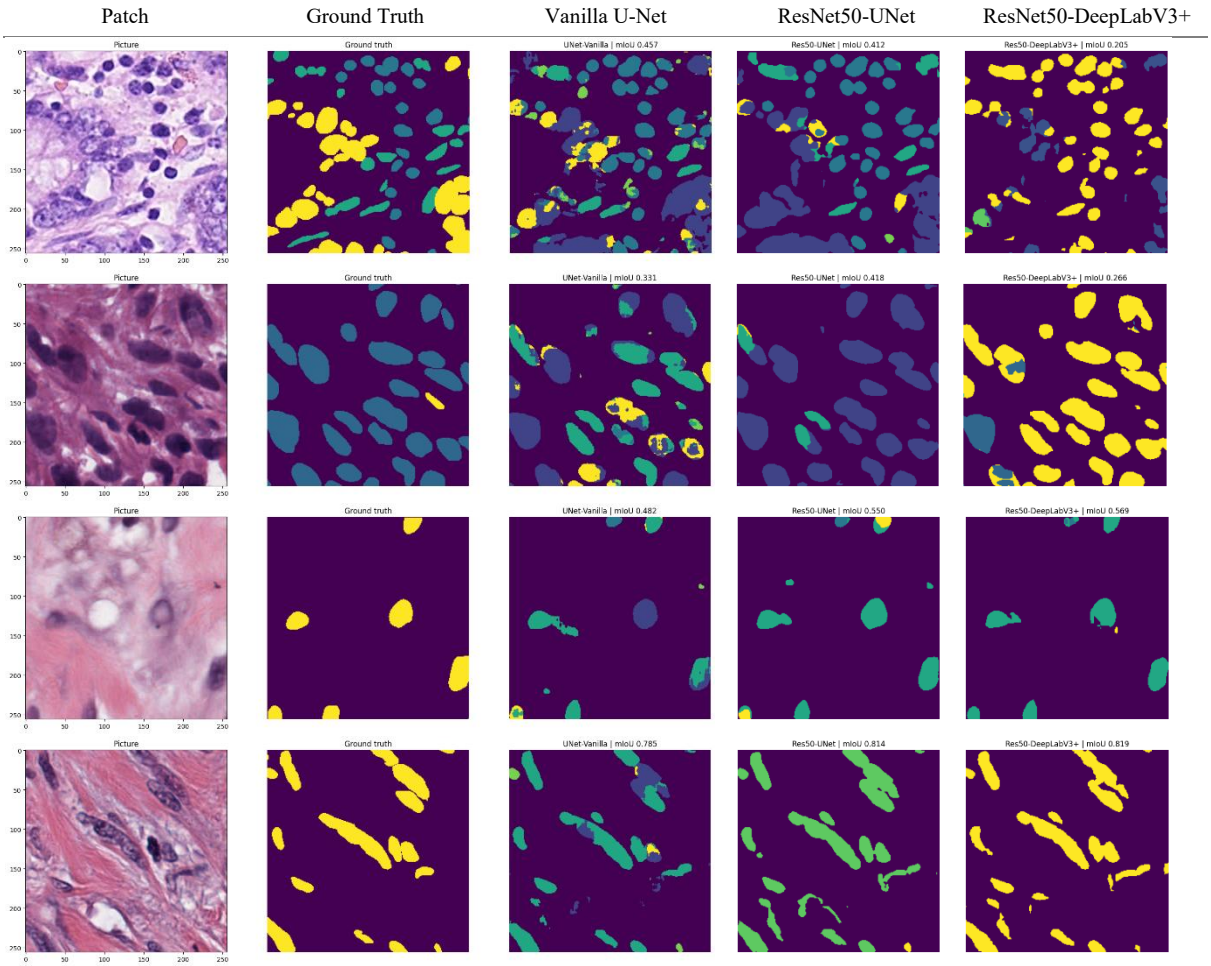
Model	Loss	Pixel Accuracy	Precision	Recall	mIoU	Dice
Vanilla U-Net	0.378	0.903	0.627	0.651	0.495	0.610
ResNet50 + U-Net	0.368	0.900	0.635	0.657	0.497	0.607
ResNet50 + DeepLabV3+	0.374	0.900	0.686	0.599	0.484	0.592

Test

Model	Pixel Accuracy	mIoU	Dice
Vanilla U-Net	0.895	0.587	0.677
ResNet50 + U-Net	0.902	0.574	0.655
ResNet50 + DeepLabV3+	0.888	0.523	0.597

Based on the evaluation, the simple U-Net architecture offered the best performance, achieving the highest mIoU and Dice scores in the control set. Although the models with the pre-trained ResNet50 encoder (ResNet50-UNet and ResNet50-DeepLabV3+) were trained significantly faster, they did not generalize as well in the control set. ResNet50-UNet provided the highest pixel accuracy, but lagged behind the most critical metrics, mIoU and Dice, based on the region. ResNet50-DeepLabV3+ showed the lowest performance across all key metrics.

Visual Results



The models showed an impressive ability to segment nuclei from the background, with the masks produced effectively detecting these structures. Cases of misidentified microscopic structures were minimal. However, a significant limitation was observed in their ability to classify nuclei. This poor performance is mainly due to the morphological diversity between the nuclei, where many share similar characteristics but belong to different categories. Although there have been rare successful classifications, such as the performance of the U-Net models in the first image and the DeepLabV3+ in the fourth, the overall classification accuracy is currently insufficient for practical application in medical imaging. Despite this, the models look promising. Above developments could be achieved through strategies such as further hyper-tuning or applying different types of codecs to more accurately identify the complex microscopic structures and patterns associated with cell nuclei.

Bibliography

1. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). MobileNetV2: Inverted Residuals and Linear Bottlenecks. ArXiv.org. https://arxiv.org/abs/1801.04381
2. Shahoveisi, F., Gorji, H. T., Shahabi, S., Hosseinirad, S., Markell, S., & Vasefi, F. (2023). Application of image processing and transfer learning for the detection of rust disease. Scientific Reports, 13(1). https://doi.org/10.1038/s41598-023-31942-9
3. Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1905.11946
4. Su, C., & Wang, W. (2020). Concrete cracks detection using convolutional NeuralNetwork based on transfer learning. Mathematical Problems in Engineering, 2020, 1–10. https://doi.org/10.1155/2020/7240129
5. Gamper, J., Koohbanani, N. A., Benes, K., Graham, S., Jahanifar, M., Khurram, S. A., Azam, A., Hewitt, K., & Rajpoot, N. (2020). PanNuke Dataset Extension, Insights and Baselines. ArXiv:2003.10778 [Cs, Eess, Q-Bio]. https://arxiv.org/abs/2003.10778
6. Cui, Z., Yao, J., Zeng, L., Yang, J., Liu, W., & Wang, X. (2024). LKCell: Efficient Cell Nuclei Instance Segmentation with Large Convolution Kernels. ArXiv.org. https://arxiv.org/abs/2407.18054v1
7. Tan, Song, B., Trinh, Kim, K., & Kwak, J. T. (2022). SONNET: A Self-Guided Ordinal Regression Neural Network for Segmentation and Classification of Nuclei in Large-Scale Multi-Tissue Histology Images. IEEE Journal of Biomedical and Health Informatics, 26(7), 3218–3228. https://doi.org/10.1109/jbhi.2022.3149936
8. Jaafar, R., Hedi Yazid, Farhat, W., & Amara, N. B. (2025). SBC-UNet3+: Classification of Nuclei in Histology Imaging Based on Multi Branch UNET3+ Segmentation Model. Proceedings of the 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, 601–609. https://doi.org/10.5220/0013232900003912
9. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI).
10. Polat, H. (2022). A modified DeepLabV3+ based semantic segmentation of chest computed tomography images for COVID-19 lung infections. International Journal of Imaging Systems and Technology, 32(5), 1481–1495. https://doi.org/10.1002/ima.22772
11. Sedat Metlek. (2024). CellSegUNet: an improved deep segmentation model for the cell segmentation based on UNet++ and residual UNet models. Neural Computing & Applications. https://doi.org/10.1007/s00521-023-09374-3
12. Isensee, F., Jaeger, P. F., Kohl, S. A. A., Petersen, J., & Maier-Hein, K. H. (2020). nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods, 18(2), 203–211. https://doi.org/10.1038/s41592-020-01008-z
13. Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A. L., & Zhou, Y. (2021). TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation. ArXiv:2102.04306 [Cs]. https://arxiv.org/abs/2102.04306
14. de Sousa, E. M. V., Kumar, A., Coupland, C., Vaz, T. F., Jones, W., Valcarce-Diñeiro, R., & Calaminus, S. D. J. (2024). U-Net as a deep learning-based method for platelets segmentation in microscopic images. https://doi.org/10.1101/2024.08.23.24312502
15. Jin, Q., Meng, Z., Pham, T. D., Chen, Q., Wei, L., & Su, R. (2019). DUNet: A deformable network for retinal vessel segmentation. Knowledge-Based Systems, 178, 149–162. https://doi.org/10.1016/j.knosys.2019.04.025
16. Agrawal, P., Katal, N., & Hooda, N. (2022). Segmentation and classification of brain tumor using 3D-UNet deep neural networks. International Journal of Cognitive Computing in Engineering, 3, 199–210. https://doi.org/10.1016/j.ijcce.2022.11.001

17. Jin, Q., Meng, Z., Sun, C., Cui, H., & Su, R. (2020). RA-UNet: A Hybrid Deep Attention-Aware Network to Extract Liver and Tumor in CT Scans. Frontiers in Bioengineering and Biotechnology, 8. https://doi.org/10.3389/fbioe.2020.605132
18. Nawaz, M., Nazir, T., Masood, M., Ali, F., Khan, M. A., Tariq, U., Sahar, N., & Damaševičius, R. (2022). Melanoma segmentation: A framework of improved DenseNet77 and UNET convolutional neural network. International Journal of Imaging Systems and Technology, 32(6), 2137–2153. https://doi.org/10.1002/ima.22750
19. Tommasino, C., Russo, C., Rinaldi, A. M., & Ciompi, F. (2023). HoVer-UNet: Accelerating HoVerNet with UNet-based multi-class nuclei segmentation via knowledge distillation. ArXiv.org. https://arxiv.org/abs/2311.12553v3
20. Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (n.d.). Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. https://arxiv.org/pdf/1802.02611
21. Long Qing Chen, Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2016). DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. ArXiv (Cornell University). https://doi.org/10.48550/arxiv.1606.00915
22. Li, K., Zhang, G., & Zhang, Y. (2024). Cell Image Segmentation Based on Improved DeepLabv3+. Proceedings of the 2024 4th International Conference on Internet of Things and Machine Learning, 195–199. https://doi.org/10.1145/3697467.3697639
23. Ahana Roy Choudhury, Rami Vanguri, Sachin Jambawalikar, & Kumar, P. (2019). Segmentation of Brain Tumors Using DeepLabv3+. 154–167. https://doi.org/10.1007/978-3-030-11726-9_14
24. Ahmad, I., Amin, J., Muhammad IkramUllah Lali, Abbas, F., & Muhammad Imran Sharif. (2024). A novel Deeplabv3+ and vision-based transformer model for segmentation and classification of skin lesions. Biomedical Signal Processing and Control, 92, 106084–106084. https://doi.org/10.1016/j.bspc.2024.106084
25. Hadinata, P. N., Djoni Simanta, Eddy, L., & Nagai, K. (2023). Multiclass Segmentation of Concrete Surface Damages Using U-Net and DeepLabV3+. *Applied Sciences*, *13*(4), 2398–2398. https://doi.org/10.3390/app13042398
26. Su, R., Zhang, D., Liu, J., & Cheng, C. (2021). MSU-Net: Multi-Scale U-Net for 2D Medical Image Segmentation. Frontiers in Genetics, 12. https://doi.org/10.3389/fgene.2021.639930