# Detection of Glaucoma Based on Optic Disc and Macular Volume Scans with Deep Learning

*Anthony Cuturrufo, Jeehyun Hwang, Daniel Israel, Shufan Li*

## Abstract

Glaucoma, one of the main causes of permanent blindness, poses considerable challenges for accurate diagnosis. Advancements in optical coherence tomography (OCT) have significantly enhanced the capability for early detection of glaucoma by providing detailed images of ocular structures. This study presents a cutting-edge deep learning method that integrates both macular and optic disc volume scans to improve the diagnostic process for glaucoma. While conventional approaches tend to evaluate these scans separately, our findings indicate that a combined analysis markedly improves diagnostic precision. We utilized a Siamese network architecture featuring dual inputs from the macular and optic regions, employing a 4-layer ResNext50 model for each stream. The merged outputs from these streams are processed further to capture extensive and complementary information from these anatomically and functionally interconnected ocular regions. Our dataset comprised OCT scans from 1132 patients, encompassing various stages of glaucoma and control subjects. By conducting training and validation across multiple data splits, we tackled the issue of class imbalance and bolstered model robustness. Our results show a notable improvement in F1 score, area under the curve (AUC), area under the precision-recall curve (AUPRC), and sensitivity, specificity. This research highlights the significance of integrating analyses from multiple ocular regions for more precise glaucoma diagnostics, thereby paving the way for enhanced screening and management strategies in clinical practice.

## Introduction

Glaucoma, a multifaceted eye condition, often results in permanent vision impairment primarily due to optic nerve damage. As a leading cause of blindness worldwide, it poses a major public health challenge. The disease's early stages are typically asymptomatic, making early diagnosis difficult but essential to halting its progression. This complexity is compounded by the interaction of various risk factors, necessitating sophisticated and accurate detection methods. (K. Schuster et al., 2020). Optical Coherence Tomography (OCT), a non-invasive imaging technique that delivers detailed cross-sectional visuals of the retina, has become essential for diagnosing and monitoring glaucoma. This high-resolution technology offers crucial insights into retinal structures, making it indispensable in the management of this eye disease. (Bussel et al., 2014).

Deep learning, a branch of machine learning known for its capability to learn layered data representations, has demonstrated potential in improving diagnostic processes, particularly in medical image interpretation. Previous research has extensively investigated the use of deep learning methods to analyze optic nerve head (ONH) data for glaucoma detection. These models

are adept at identifying subtle disease patterns that may be missed by even seasoned clinicians. Moreover, the rise of geometric deep learning has enabled the incorporation of topological and geometric information into model evaluations, providing detailed insights into the intricate structure of the optic nerve and its changes in glaucomatous eyes (Braeu et al., 2023; George et al., 2020; Maetschke et al., 2019; Thiéry et al., 2023).

Despite these advancements, the macula, which also undergoes structural changes in glaucoma, has not been as thoroughly investigated using these novel technologies. Only a limited number of studies have delved into 3D Macular OCT scan analysis for glaucoma diagnosis, one of which meticulously labeled the segmentation of retinal layers, pointing to the potential of detailed macular data in understanding the disease's pathology (Russakoff et al., 2020). Another more recently, proving the effectiveness of 2D and 3D models for glaucoma diagnoses as well (Rasel et al., 2024).

This research seeks to address the diagnostic challenge by combining optic nerve and macular volume scans, utilizing deep learning to extract complementary information from each region. Through the use of a dual-input Siamese network architecture to analyze these interconnected areas, we expect our model to offer a more comprehensive and precise representation of glaucoma. This approach aims to enhance diagnostic accuracy, facilitating earlier and more dependable detection of the disease.

This paper details our methodology and results, representing a significant advancement in utilizing OCT scans for glaucoma diagnosis. We believe this approach can act as a foundational framework for future studies, particularly in applying deep learning to extensive ocular imaging.

# Methods

## Dataset Collection

The dataset utilized in this study was sourced in collaboration with UCLA Health, encompassing a private collection of optical coherence tomography (OCT) scans. This dataset includes both macular and optic disc volume scans from a diverse patient population. The OCT scans were collected over a period of five years, from 2015 to 2020. Patients ranged in age from 18 to 85 and were recruited based on routine clinical evaluations or glaucoma screenings. Each participant provided informed consent prior to the inclusion of their data in the study. The diagnosis of glaucoma was established through a comprehensive clinical evaluation, which included intraocular pressure measurement, visual field testing, and a thorough examination of the optic nerve head by experienced ophthalmologists. Each diagnosis was corroborated by a glaucoma specialist to ensure accuracy. Patients were categorized into either or as normal controls based on established clinical criteria. To maintain patient confidentiality, all identifiable information was removed from the dataset. Each scan was assigned a unique identifier that does not trace back to any personal patient information.

## Study Sample

The dataset contains both Macular and Optic Disc volume scans for each patient. The dataset was approximately divided into training, validation, and testing groups at ratios of 80%, 7%, and 13% respectively, based on the total number of patients in the dataset. Specifically, 150 patients were selected for the test set, including some whose scans consistently showed either all

glaucoma or all normal, and others who initially had normal scans that later developed signs of glaucoma.

Test Set

| Macular Volume Scans | Optic Nerve Volume Scans |
|---|---|
| 130 Patients With Glaucoma Scans<br>26 Patients With Normal Scans<br>6 Patients Having Both Normal and Glaucomatous Scans | |
| 221 Glaucoma Eyes<br>48 Normal Eyes | 226 Glaucoma Eyes<br>52 Normal Eyes |

*Figure 1: Test Set Summary. The 'eyes' row represents the count of individual eyes (left or right) diagnosed with either glaucoma or normal conditions.*

The remaining patients were divided into five training/validation splits for model selection and hyperparameter tuning. To compensate for the limited non-glaucoma data, we augmented the negative class of the dataset with any previously unlabeled scans from a patient taken before their most recent negative glaucoma diagnosis. 214 normal scans from the macular region and 556 normal scans from the optic nerve region were added to the training set from this augmentation method. Each validation split was created by holding out an additional randomly selected 75 patients (≈10% of all patients). Note that the augmentation method did not add any scans to the validation set. The rest of the scans were used for training.

Training + Validation Set

| Macular Volume Scans | Optic Nerve Volume Scans |
|---|---|
| 853 Glaucoma Patients<br>123 Normal Patients<br>41 Patients Having Both Normal and Glaucomatous Scans | |
| 1503 Glaucoma Eyes<br>219 Normal Eyes | 1320 Glaucoma Eyes<br>227 Normal Eyes |
| 1505 Glaucoma Scans<br>433 Normal Scans | 1322 Glaucoma Scans<br>830 Normal Scans |

*Figure 2: Training and Validation set summary. The 'scan' row denotes the total number of individual Macular and Optic Nerve Volume scans conducted on patients diagnosed with either glaucoma or normal eye conditions. Each entry in these rows tallies the total scans, acknowledging that some patients may have had multiple scans. This accounts for the number of scans being higher than the count of eyes, reflecting repeat examinations or follow-ups.*

The test set demonstrated a patient class imbalance of approximately 0.167, indicated by the ratio of normal to glaucomatous patients. In the training and validation set, this imbalance was slightly lower at approximately 0.126. To address these imbalances, we introduced a weighted loss of 0.15 within the loss function, which serves to mitigate the impact of class imbalance on model performance. Further details on the implementation of the weighted loss and its effects on model learning will be elaborated in the Model Specifications section.

## OCT Imaging

The crux of our methodology relies on the precision and quality of the imaging modalities employed. In our study, we utilized a CIRRUS™ HD-OCT machine, known for its high-definition imaging capabilities and widespread clinical adoption. The CIRRUS™ platform

generates detailed cross-sectional images of the retina, which are paramount for detecting and quantifying the subtle anatomical changes associated with glaucoma.

For the purpose of our deep learning model, we used the raw volumetric OCT scan data. This approach allowed us to preserve the rich, intricate details present in the original images, which are often lost or compressed in processed or summarized formats. By training and evaluating our models on these raw scans, we aimed to capitalize on the full spectrum of information inherent in the images, ensuring that our deep learning algorithms could learn to recognize the most nuanced indicators of pathology.

The OCT scans encompass both macular and optic nerve head (ONH) regions, providing a comprehensive view of the ocular structures. Our decision to employ raw images directly parallels the real-world scenario where clinicians assess OCT images in their unaltered form, fostering a level of diagnostic precision in our model that mirrors clinical practice (Sung et al., 2012). This alignment with clinical evaluation techniques is critical in the translational applicability of our research, enabling a smoother integration into clinical workflows and aiding in the acceleration of its adoption in real-world diagnostic settings.

## Data Processing

We optimized the original Cirrus OCT volume scans initially sized at 1024x200x200 pixels, by downsizing them to 128x200x200 pixels. This approach optimizes the memory and runtime of the training process and reduces the risk of overfitting. In addition, we enhanced the training sets with various data augmentations including a composition of random Gaussian noise, pixel intensity adjustments, and affine transformations. In addition, we also added volumetric scan denoising using BM3D for its prior ability to improve classification accuracy (Chan et al., 2017). The denoising is done by first identifying and grouping similar 2D patches from the noisy image into 3D stacks, then applies a transformation to distinguish and suppress the noise more effectively, and finally inversely transforming and reintegrating the filtered patches back into the original image (Lebrun, 2012).

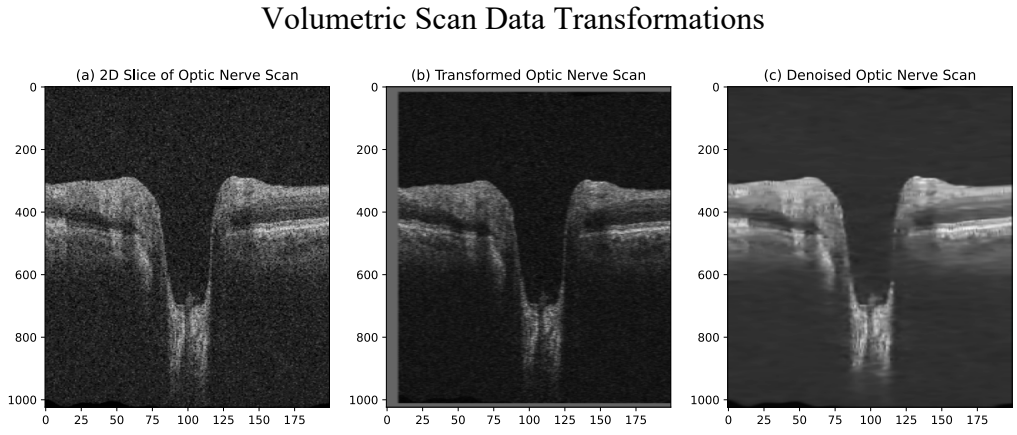


*Figure 3: Volume Scan Transformations: (a) a 2D Slice from original Optic Nerve Scan Dataset. (b) shows an augmentation technique to diversify the dataset. There is pixel intensity adjustment making it appear darker along with a slight affine transformation shifting it to the bottom-right corner. (c) displays the optic nerve scan post-denoising, where the previously*

*visible noise, especially noticeable around the top and bottom edges of the scan, has been significantly reduced. The denoising process has also sharpened the edges of the optic nerve, making the boundaries and internal structures more distinct.*

## Model Specifications

After experimenting with numerous deep image models including 3DCNN, ResNet, and ViT architectures, we selected a ResNext50 architecture (Xie et al., 2017) from the MONAI library (Cardoso et al., 2022) for its prior effectiveness with data-limited medical datasets given the relatively small size of our study sample and its outperformance of the ResNet architecture (Tharsanee et al., 2021; Xie et al., 2017). We adjusted important regularization hyperparameters such as dropout and weight decay during hyperparameter tuning. We used the Adam optimizer within a Pytorch Lightning framework. Our model employed a binary cross-entropy loss function, which included an additional weight of 0.15 applied to the positive (glaucoma) class. This weight modifies the contribution of the glaucoma class to the total loss, leading to an increased penalty for diagnosing glaucoma on a non-glaucoma scan. The weight value reflected the approximate class imbalance ratio of normal to glaucoma patients in the dataset. We set the learning rate at 0.00001. The batch size was set to 36, based on the performance on the fine-tuning runs, this was also the maximum supported batch size by the memory constraints of the GPU. All experiments were conducted using an NVIDIA L40S GPU, equipped with 46GB of memory. Each training session was conducted over 300 epochs and took about 12 hours to complete.

### Macular and Optic Disc Volume Scan Training

As the data for every patient predominantly features scans from both macular and optic disc regions, we created a model designed to analyze scans from both regions simultaneously. This dual input region approach ensures a thorough evaluation, combining scans from each region for the same patient to provide a more complete understanding of their condition, thereby improving diagnostic accuracy. A Siamese network architecture (Bertinetto et al., 2016) was selected for its previous uses in dual input stream classification (Bertinetto et al., 2016; Jia et al., 2022; Valero-Mas et al., 2024)

Architecturally, the model comprises two ResNext networks each with the same architecture as the single-scan model, one for both Macular and Optic regions. Their final convolutional layers, both of dimension 512, are concatenated. This combined output is then processed through an additional fully connected neural network, effectively harnessing the distinct yet complementary insights from both types of scans to enhance diagnostic performance.

Single and Dual-Input Macular and Optic Model

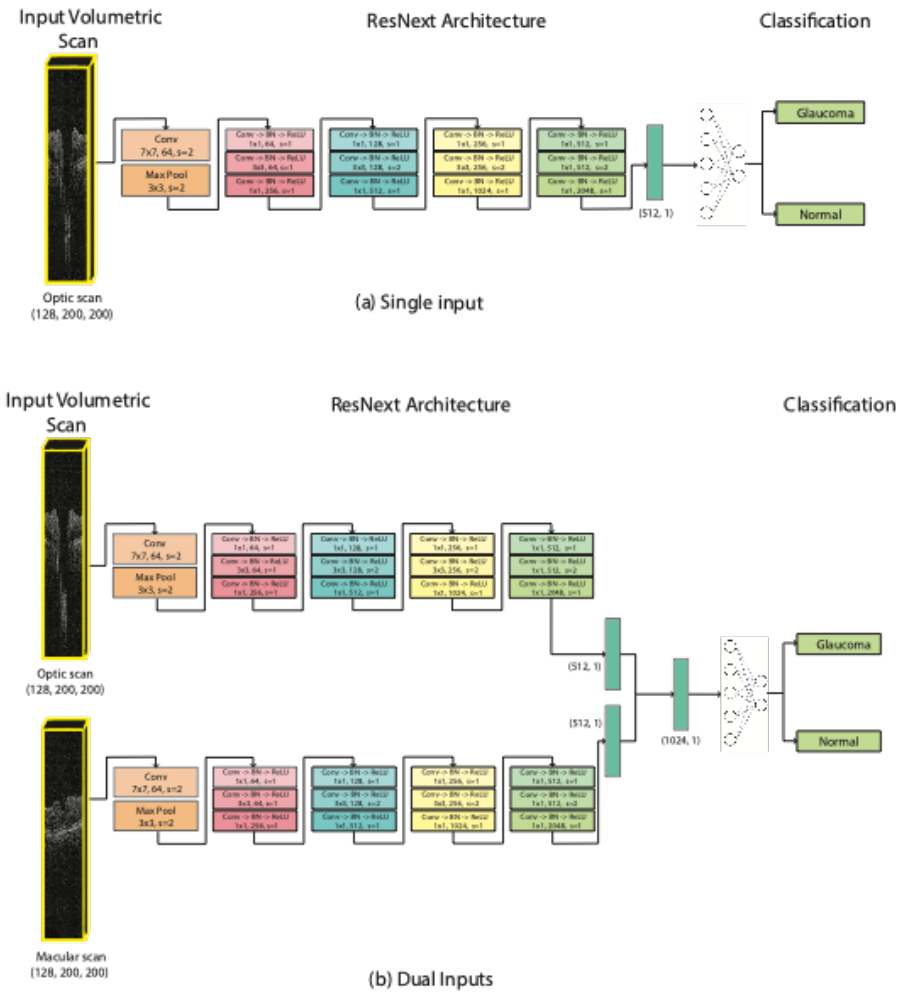


*Figure 4: Single and Dual-Input Macular and Optic Model. (a) describes the model architecture designed to process volumetric scans from either the Optic Nerve Region or the Macular region of the eye, exclusively. (b) outlines the model architecture for the dual input model, which accepts volumetric scans from both the Optic Nerve Region and the Macular region simultaneously.*

## Results

Our study analyzed a dataset in total comprising OCT volume scans from 2008 eyes of 1132 patients, including 983 with glaucoma (1729 eyes) and 149 normal control subjects (279 eyes). Table 1 provides the findings with regard to the diagnostic performance of Optic disc and macular volume scans separately and in combination. We report the mean and standard deviation of the performance metrics for each model on the test set, where models were trained using five distinct train/validation splits. For each split, we selected the epoch that achieved the highest average of sensitivity and specificity during validation. This selection criterion was chosen to address the instability of sensitivity and specificity caused by class imbalance in the dataset. The chosen model was then evaluated on the test set.

| | Macular | Optic | Macular + Optic |
|---|---|---|---|
| F1 | .8059 ± .043 | .8483 ± .039 | .8588 ± .028 |

| | | | |
|---|---|---|---|
| AUC | .8700 ± .035 | .8804 ± .030 | .9180 ± .031 |
| AUPRC | .8843 ± .011 | .8780 ± .014 | .8903 ± .010 |
| Sensitivity | .8146 ± .095 | .8155 ± .094 | .8612 ± .090 |
| Specificity | .8213 ± .092 | .8093 ± .083 | .8556 ± .075 |

*Figure 5: Diagnostic Performance Metrics of OCT Volume Scans for Glaucoma Detection. This figure illustrates the results of analyzing OCT scans for glaucoma. The study presents the mean and standard deviation of F1 score, AUC, AUPRC, sensitivity, and specificity for the detection of glaucoma using macular, optic disc, and a combined macular plus optic disc scans. Models were optimized across five training and validation splits to mitigate the effects of class imbalance and ensure stability of the diagnostic metrics.*

## Discussion

The implementation of a deep learning model utilizing both macular and optic disc scans has demonstrated significant promise in glaucoma detection. This integrated analysis resulted in marked improvements in critical diagnostic metrics such as F1 score, AUC, and sensitivity. Specifically, the combined approach exhibited an increase in the AUC to 0.9180 ± 0.031 and an F1 score of 0.8588 ± 0.028, with sensitivity rising to 0.8612 ± 0.090. Although there was a slight increase in specificity variability, the reduced standard deviations across all metrics highlight the model's consistent performance.

A primary limitation of this study is the relatively small and imbalanced dataset, which could affect the generalizability of the results. While the enhanced performance metrics and low variability demonstrate the model's robustness, the minor increase in specificity variability indicates a need for a more extensive dataset to achieve balanced sensitivity and specificity. Future research will focus on expanding the dataset to include a larger, more representative sample of OCT scans, potentially enhancing the model's reliability and applicability in clinical settings. Overall, while the findings are promising and underscore the model's strong diagnostic capabilities, they also emphasize the importance of broader datasets to address class imbalance and further refine model accuracy, particularly in a clinical context where high sensitivity and specificity are critical.

## References

Bertinetto, L., Valmadre, J., Henriques, J. F., Vedaldi, A., & Torr, P. H. S. (2016). Fully-Convolutional Siamese Networks for Object Tracking. In G. Hua & H. Jégou (Eds.), *Computer Vision – ECCV 2016 Workshops* (pp. 850–865). Springer International Publishing. https://doi.org/10.1007/978-3-319-48881-3_56

Braeu, F. A., Thiéry, A. H., Tun, T. A., Kadziauskiene, A., Barbastathis, G., Aung, T., & Girard, M. J. A. (2023). Geometric Deep Learning to Identify the Critical 3D Structural Features

of the Optic Nerve Head for Glaucoma Diagnosis. *American Journal of Ophthalmology*, *250*, 38–48. https://doi.org/10.1016/j.ajo.2023.01.008

Bussel, I. I., Wollstein, G., & Schuman, J. S. (2014). OCT for glaucoma diagnosis, screening and detection of glaucoma progression. *British Journal of Ophthalmology*, *98*(Suppl 2), ii15–ii19. https://doi.org/10.1136/bjophthalmol-2013-304326

Cardoso, M. J., Li, W., Brown, R., Ma, N., Kerfoot, E., Wang, Y., Murrey, B., Myronenko, A., Zhao, C., Yang, D., Nath, V., He, Y., Xu, Z., Hatamizadeh, A., Myronenko, A., Zhu, W., Liu, Y., Zheng, M., Tang, Y., … Feng, A. (2022). *MONAI: An open-source framework for deep learning in healthcare* (arXiv:2211.02701). arXiv. https://doi.org/10.48550/arXiv.2211.02701

Chan, G. C. Y., Muhammad, A., Shah, S. A. A., Tang, T. B., Lu, C.-K., & Meriaudeau, F. (2017). Transfer learning for Diabetic Macular Edema (DME) detection on Optical Coherence Tomography (OCT) images. *2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, 493–496. https://doi.org/10.1109/ICSIPA.2017.8120662

George, Y., Antony, B. J., Ishikawa, H., Wollstein, G., Schuman, J. S., & Garnavi, R. (2020). Attention-Guided 3D-CNN Framework for Glaucoma Detection and Structural-Functional Association Using Volumetric Images. *IEEE Journal of Biomedical and Health Informatics*, *24*(12), 3421–3430. https://doi.org/10.1109/JBHI.2020.3001019

Jia, S., Jiang, S., Lin, Z., Xu, M., Sun, W., Huang, Q., Zhu, J., & Jia, X. (2022). A Semisupervised Siamese Network for Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 1–17. https://doi.org/10.1109/TGRS.2021.3116138

K. Schuster, A., Erb, C., M. Hoffmann, E., Dietlein, T., & Pfeiffer, N. (2020). The Diagnosis and Treatment of Glaucoma. *Deutsches Ärzteblatt International*, *117*(13), 225–234. https://doi.org/10.3238/arztebl.2020.0225

Lebrun, M. (2012). An Analysis and Implementation of the BM3D Image Denoising Method. *Image Processing On Line*, *2*, 175–213. https://doi.org/10.5201/ipol.2012.l-bm3d

Maetschke, S., Antony, B., Ishikawa, H., Wollstein, G., Schuman, J., & Garnavi, R. (2019). A feature agnostic approach for glaucoma detection in OCT volumes. *PLOS ONE*, *14*(7), e0219126. https://doi.org/10.1371/journal.pone.0219126

Rasel, R. K., Wu, F., Chiariglione, M., Choi, S. S., Doble, N., & Gao, X. R. (2024). Assessing the efficacy of 2D and 3D CNN algorithms in OCT-based glaucoma detection. *Scientific Reports*, *14*(1), 11758. https://doi.org/10.1038/s41598-024-62411-6

Russakoff, D. B., Mannil, S. S., Oakley, J. D., Ran, A. R., Cheung, C. Y., Dasari, S., Riyazzuddin, M., Nagaraj, S., Rao, H. L., Chang, D., & Chang, R. T. (2020). A 3D Deep Learning System for Detecting Referable Glaucoma Using Full OCT Macular Cube Scans. *Translational Vision Science & Technology*, *9*(2), 12. https://doi.org/10.1167/tvst.9.2.12

Sung, K. R., Wollstein, G., Kim, N. R., Na, J. H., Nevins, J. E., Kim, C. Y., & Schuman, J. S. (2012). Macular assessment using optical coherence tomography for glaucoma diagnosis. *British Journal of Ophthalmology*, *96*(12), 1452–1455. https://doi.org/10.1136/bjophthalmol-2012-301845

Tharsanee, R. M., Soundariya, R. S., Kumar, A. S., Karthiga, M., & Sountharrajan, S. (2021). 7—Deep convolutional neural network–based image classification for COVID-19 diagnosis. In U. Kose, D. Gupta, V. H. C. de Albuquerque, & A. Khanna (Eds.), *Data*

*Science for COVID-19* (pp. 117–145). Academic Press. https://doi.org/10.1016/B978-0-12-824536-1.00012-5

Thiéry, A. H., Braeu, F., Tun, T. A., Aung, T., & Girard, M. J. A. (2023). Medical Application of Geometric Deep Learning for the Diagnosis of Glaucoma. *Translational Vision Science & Technology*, *12*(2), 23. https://doi.org/10.1167/tvst.12.2.23

Valero-Mas, J. J., Gallego, A. J., & Rico-Juan, J. R. (2024). An overview of ensemble and feature learning in few-shot image classification using siamese networks. *Multimedia Tools and Applications*, *83*(7), 19929–19952. https://doi.org/10.1007/s11042-023-15607-3

Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. (2017). *Aggregated Residual Transformations for Deep Neural Networks* (arXiv:1611.05431). arXiv. https://doi.org/10.48550/arXiv.1611.05431