Estimating Intracranial Pressure Using OCT Scans of the Eyeball

Eric He
Center for Data Science
New York University
New York, NY 10011
eh1885@nyu.edu

Farris Atif
Center for Data Science
New York University
New York, NY 10011
fda239@nyu.edu

Gadi Wollstein, MD
Grossman School of Medicine
New York University
New York, NY 10011
gadi.wollstein@nyulangone.org

Nasser Al-Rayes
Center for Data Science
New York University
New York, NY 10011
nka8061@nyu.edu

Zixiao Chen
Center for Data Science
New York University
New York, NY 10011
zc2194@nyu.edu

Narges Razavian, PhD
Grossman School of Medicine
New York University
New York, NY 10011
narges.razavian@nyulangone.org

Abstract

Convolutional neural networks (CNNs) have gained great importance in the space of computer vision with large applications in the medical space. We fit 3D CNNs on optical coherence tomography (OCT) scans of the eyeball to measure the relationship between an organism's eyeball shape, intraocular pressure ("IOP") and intracranial pressure ("ICP"). Since we only had around 1500 scans to work with, we attempted various training techniques to solve data scarcity problems, including data augmentation, warm-starting from pretrained models and self-supervised learning methods. Because we had to spend a lot of time consolidating, evaluating, and transforming the data into a model-ready state, we were unable to fully train models to convergence, or explore tasks like interpreting the model. We will continue experimenting with different hyperparameter settings, other pretraining methods, and larger model architectures over the course of the next semester. You can find our code at https://github.com/Bulbasaurzc/Capstone2021.

1 Introduction

1.1 Intracranial pressure, intraocular pressure, and the lamina cribrosa

ICP is the fluid pressure between the skull and brain tissue. Anomalous ICP values are associated with health issues including headaches, brain swelling, blindness, and tumors. Though knowing an organism's ICP can be useful in clinical settings, directly measuring it is a dangerous and invasive procedure. In this research, we prototype a way to indirectly estimate ICP via its relationship with intraocular pressure ("IOP"), the fluid pressure inside the eye, and the lamina cribrosa (LC), a connective tissue network located at the back of the eye.

The LC serves as structural and nutritional support to the retinal ganglion cell axons which transmit signals from the eye to the brain, and its deformation is understood to be a cause of glaucoma and other blinding diseases. As the LC is a 3D meshwork made up of collagen fibers, its shape can be deformed by local pressures in difficult-to-predict ways. The IOP exerts itself on the LC from the anterior dimension, while the ICP acts on the posterior and centripetal aspects. The IOP and ICP

Submitted to 35th Conference on Neural Information Processing Systems (NeurIPS 2021). Do not distribute.

are not directly opposing each other, so the interplay of the two pressures can result in complex deformations of the LC. Due to the difficulty of measuring ICP, however, the precise relationship between ICP, IOP, and LC structure remains poorly understood.

In [14], 14 rhesus monkeys were anesthetized, and their ICPs and IOPs were manipulated via cannulation of the eye and lateral ventricle in the brain. Scans of the optic nerve head and the LC within it were taken in vivo during these manipulations using optical coherence tomography, a technology commonly used in ophthalmology to produce in vivo, 3D, micron-resolution scans. We thus have a 4TB dataset of LC scans where the ICP and IOP values are known.

In this paper, we use the monkey dataset to train machine learning models to predict ICP given the IOP value and the LC structure. Because an organism's IOP and LC can be measured in real time and non-invasively, an accurate model would be a first step towards a safe and easy method to estimating the ICP.

2 Related Work

Deep learning methods have seen successes in many fields, including healthcare [8], self-driving [11], chat-bots [16], and across many data modalities such as audio [5], vision [13], and text [17]. We train a deep vision model on our dataset to see if it can accurately predict ICP given IOP and the OCT scan. However, there are some differences relative to the standard image modeling task:

1. The OCT is a sizable 3D image, whereas standard CV approaches assume 2D images. Examples of deep learning models for 3D image are relatively abundant in the medical field, however; [1] is a repository of 3-D ResNets [6], while [7] and [3] trained 3D models to detect OCT scans of glaucoma in the eye.
2. Most deep learning models are tuned for classification tasks, but this task is a regression problem.
3. In addition to the OCT scan, it is critical to consider the IOP value in our ICP prediction, as one LC structure can be associated with many different combinations of IOP and ICP. This can be considered a simple form of multimodal learning.
4. We do not have a large amount of data to work with, which can make training a model from scratch quite difficult. Large 2D image models [10] are typically trained on extremely large datasets but no analogue exists for 3D medical scans. Transfer learning is a standard approach to low-data situations in which a model already trained on other datasets is fine-tuned for the low-data task. However, we have found papers that suggests it is difficult for many pretrained models to generalize to the domain of medical imaging data.
5. As a result, we also evaluate self-supervised learning methods for pretraining our image models. Self-supervised learning starts model training on proxy tasks designed to force the model to learn a representation of the data; for example, sections of the image could be blurred and the model would have to reconstruct the missing data. Various self-supervised tasks exist [9] [4]. [12] evaluates a number of self-supervised learning tasks in the context of medical imaging.

Apart from the medical value of being able to measure ICP non-invasively, we hope our work on 3D computer vision, deep regression, multimodal learning, and self-supervised pretraining can be of reference to the machine learning community.

3 Problem Definition and Algorithm

3.1 Task

Our goal is to accurately predict ICP using properties of the eye - IOP and the resulting deformations of the lamina. These deformations are believed to be detectable using high-resolution OCT scans of

the eye. Due to the resolution of these scans, small perturbations in the IOP are believed to cause structural lamina changes, as the pressure exerted on this tissue change. Such changes in pressure should cause its shape to bend.

3.2 Algorithm

Our belief is a well-trained CNN would detect such structural deformations, and when given an objective (in this case, ICP), would learn how changes in the ICP would impact the lamina. Thus, when presented with scans of a lamina, a trained CNN would be able to accurately predict a patient's ICP, removing the need to perform any invasive brain-pressure measuring procedures.

4 Experimental Evaluation

4.1 Data

Much of the time spent this semester was spent preparing the dataset rather than training on it. We summarise some of our decisions in preparing the data for reference.

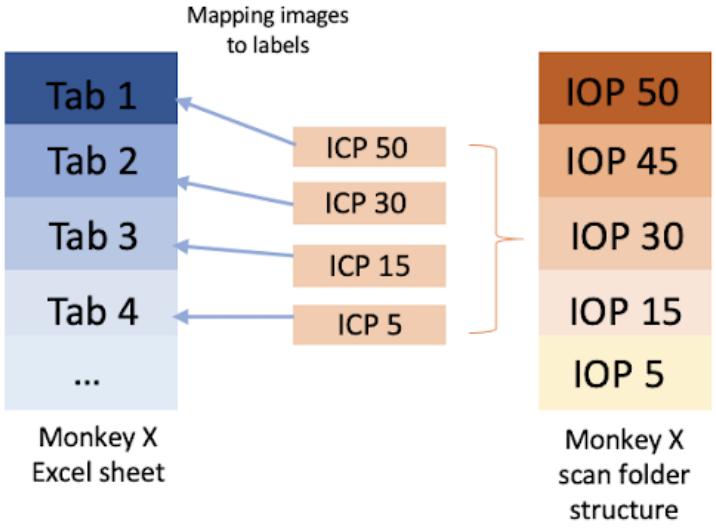

Figure 1: Data processing flow: Each tab in a monkey Excel sheet maps to one monkey scan sub-folder that represents a certain ICP/IOP combination.

The monkey data was gathered over the course of 6 years, with different sets of monkeys having their data organized according to different file conventions depending on the lab assistant in charge at the time. For example, monkeys 1-5 have a hierarchical folder structure where the file paths are inferred from the IOP and ICP values in an Excel sheet (or vice versa), while monkeys 11-14 dump all scans into one directory and attach a "scan ID". We manually mapped all info into a master dataset which holds a new "ID" we created to identify each scan, IOP and ICP values where we could infer them, and the file path to the original scan data.

Most IOP and ICP values were not filled out for the following reasons:

1. Some of these were intentional baseline readings taken before the IOP/ICP were altered. We excluded these from the regression training set, but still use them to perform self-supervised learning.
2. The vast majority of scans did not have a reading taken at time of scan; for these, we generally used the last known IOP/ICP values.

3. A minority could not be inferred from the filepath or had values outside the parameter range of interest. For example, monkey 10 has many scans with negative ICP values taken after it became deceased. We removed these from any training sets.

At the end, we still had 327 scans with missing IOP values and 297 scans with missing ICP values which could not be used for training.

The scan dimensions could be very different depending on the OCT machine as well as the type of scan being taken. For example, monkeys 10-14 primarily had scans of shape [400,400,1024] (height, width, depth) while monkeys 1-6 primarily had scans of shape [512, 512, 2048]; other scans focused on very specific areas of the eyeball and could have shapes like [600, 100, 1024]. We kept all scans where the height and the width were equivalent, and the depth was larger than both. A table of scan dimensions is shown below; we were left with 1486 usable scans.

Table 1: Scan dimensions.

Part		
Scan Size	Count	Eligible for training
400, 400, 1024	799	Yes
512, 512, 2048	687	Yes
600, 100, 1024	84	No
600, 100, 1024	26	No
400, 300, 1024	9	No
100, 1000, 1024	3	No
300, 230, 1024	2	No
120, 184, 1024	2	No

Our final preprocessing step standardizes the images to have mean 0 and variance 1 to improve model stability. During training, we apply standard data augmentations: flips, affine transforms, blurs, and Gaussian noise.

4.2 Methodology

Given IOP and 3D image scans, our goal is to predict ICP. The prediction flow is shown in the figure above. The fundamental building block is the 3D resnet [15] which synthesizes the scan data into interpretable features for a feedforward network. The scan data is then fed through a 3D ResNet and squashed through a linear layer to output a set of "image features". At this point, the IOP data is concatenated with the image features. A single IOP value is not given much weight by the model, however, so we broadcast the scan's IOP value into a long tensor for concatenation. The concatenated IOP tensor and image feature tensor is fed into a multi-layer perceptron (MLP) which outputs the final prediction, a standardized ICP value. The flow is shown in Figure 2.

4.2.1 Warm Starting from a Pretrained Model

The performance of deep learning models is greatly impacted by the quantity of training data, and we only have 1500 scans from 14 monkeys. We could not guarantee that we could sufficiently train any model from scratch, and finding the optimal hyperparameters would be even harder. Taking a pretrained model and finetuning it on our task could greatly increase the likelihood and speed of training convergence as well as improving the final model accuracy.

We used Tencent's MedicalNet [1] as another candidate model. The MedicalNet is a set of ResNets [6] which were pretrained on MRI scan data of the brain and then trained again on 3D scans of the monkey's eyeballs. Doing so would allow us to see how a pretrained model would perform when compared to a model that is trained from scratch.

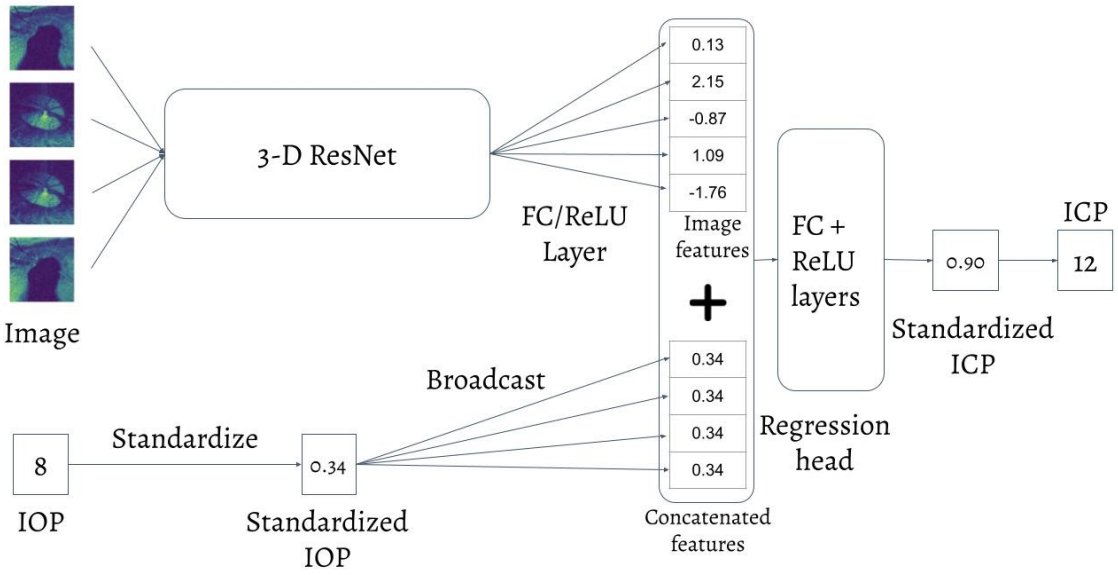

Figure 2: Training a ResNet from scratch and concatinating tabular data into the first fully-connected layer.

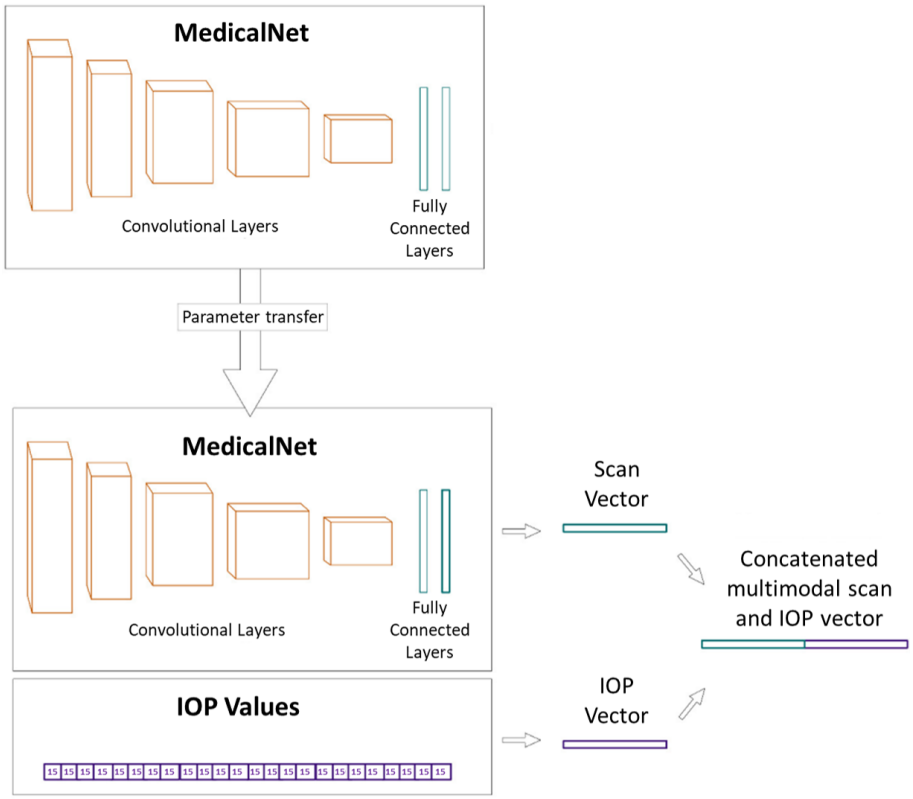

Figure 3: Transferring hyperparameters from a pretrained MedicalNet model to the MedicalNet model used for our purposes. We used the hyperparameters from the pretraining and also inject tabular IOP values with the hope of speeding up training convergence.

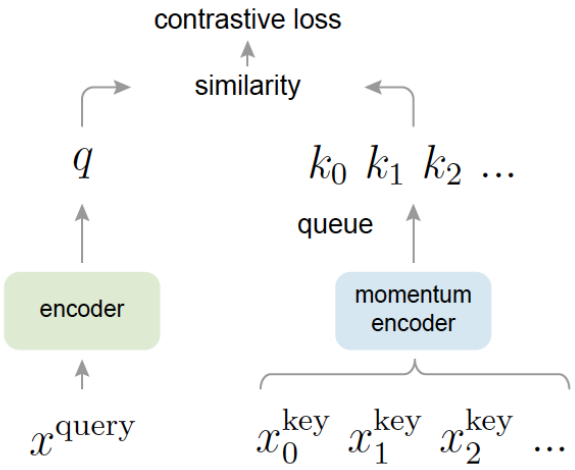

Figure 4: The original Momentum Contrast illustration from [2].

4.2.2 Self-Supervised Learning

Another pretraining method, applicable to both models initialized from scratch and pretrained models, is self-supervised learning. In self-supervised learning, a model is first trained on various "pretext" tasks with the goal of forcing the model to learn a better internal representation of the images. Our pretext task is to classify whether two random crops came from the same original image.

We attempted the momentum contrastive loss ("MoCo") method [2] to pretrain models. The Resnet model is designated the "encoder", and is asked to produce an embedding for one image crop. A second "decoder" model, whose parameters are an average of the historical encoder weights, embeds a second image crop. The dot product of the two embedding tensors is fed through a sigmoid to determine the predicted probability the crops come from the same image.

4.3 Results

The large scale of the data as well as hardware constraints posed many challenges during modeling. As a result, batch sizes were kept relatively small (8-16) across different architectures. While initial models worked well to minimize training loss, the results did not generalize to holdout sets. To offset this, the input data was perturbed (bias was added), which significantly improved results when passed through ResNet50. Additionally, the Adamax optimizer (lr 3e-4) resulted in better convergence than that of the out of box MedNet SGD. The figures below summarize the results:

Table 2: Model configuration and results.

Model	Augmented	Epoch	Pretrained	Validation Loss
ResNet50	Yes	20	No	.65
ResNet50	No	20	Yes	.43
ResNet 10	Yes	100	No	.39

While ResNet10 minimized mean validation loss across epochs, closer analysis of the learning curves reveals model configuration 1 (ResNet50 w/ Data Augmentation) yields promising results. It should be mentioned that albeit the lack of convergence in the pre-trained models, an extensive grid search over network parameters may yield results which successfully utilize the "warm-start". Moreover, when the holdout set was created by shuffling images regardless of monkey assignment - these models began to minimize validation loss.

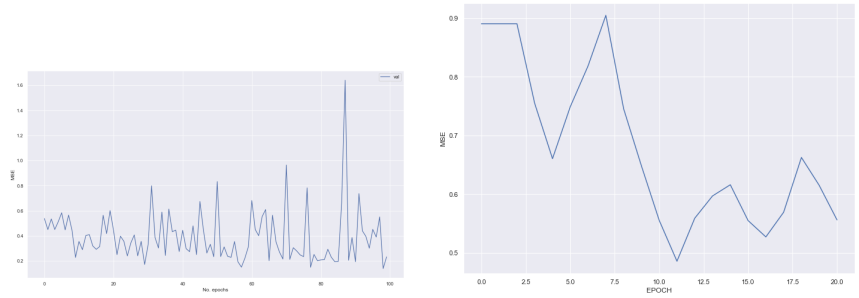

Figure 5: Validation set learning curves for ResNet10 (left) and ResNet50 (right).

4.4 Discussion

With a limited amount of time, we were able to restructure the large dataset and map numeric data to image scans. We also managed to build a multi-modal learning architecture by leveraging deep learning models. However, there still are lots of potential methods which we could implement and experiment with:

1. **Change the way we predict IOP** Instead of concatenating the IOP value into the network, we could have changed the prediction task to predict the ratio of the ICP to IOP; this would have obviated the need to include additional linear layers after the concatenation at the end of the network.
2. **Try bigger models** Smaller models like ResNet10 could only predict the mean, but we were able to get ResNet50 to continuously lower validation loss. Bigger models and other hyperparameters may do a better job.
3. **Try more pretrained models** MedNet was trained on MRI data, which is very different from our OCT data. We could not find recently OCT-pretrained models, but training one ourselves on larger datasets before using it for ICP prediction may have yielded a better performing model.
4. **Solving issues with distributed training** We were prevented from using MoCo only due to technical difficulties around distributed training. With such sparse data, a successfully SSL-trained model could have gotten much better results.
5. **Model explanations** The ResNet50 validation loss is still dropping; if it gave reasonably accurate predictions, explanatory methods could help understand the model better.

5 Conclusion

We attempted to train deep neural networks on 3D eyeball medical scans, and were unable to attain the level of validation convergence needed to prove that model learning was occurring. Given the quantity of time spent data cleaning, we made great strides in modeling nonetheless. Our initial iterations on both the trained-from-scratch and pretrained models showed convergence with the training set, and even slight convergence with the validation set. We believe that with further hyperparameter tuning and model architecture adjustments, one would be able to attain validation convergence.

6 Lessons Learned

From this project we learnt that data sets provided in real-world data science problems are much larger, messier, and more complicated than canned data sets. We also learnt that pre-training and self-supervised learning could be useful techniques to apply when dealing with relatively small data sets, especially when adding new data is difficult due to humanitarian concerns (such as animal experiments).

7 Student Contributions

Eric worked on exploring and munging the data, gathered all the data into Google Drive and converted the data into PyTorch, collaborated with Farris on writing code to train, checkpoint, and evaluate models from scratch, worked on self-supervised learning, and was primary writer of the Intro/Related Work/Data sections of the paper.

Farris created the infrastructure on Greene to train the ResNet architectures (ie. creating a pipeline from Google Drive to Greene to transfer data, package management etc.). Worked with Nasser and Eric to modify pre-trained and from-scratch models for tabular data injection and bash shell execution. Primary author of the Results section.

Nasser worked on proving the viability of using the MedicalNet models with our data. He conducted preliminary tests with the models from this repository and showed that the models could be trained on our data. He also modified the architecture for it to return a numerical, regression output and ran training jobs on pretrained models with the data.

Zixiao worked on initial data issues structuring, data cleaning and mapping numerical data to image data. Worked on data loader building and exploratory data analysis. Collaborated with Eric on writing Data section of this paper.

7.1 Acknowledgements

We would like to thank our mentor Dr. Gadi Wollstein, Dr. Narges Razavian, and our lab manager Ronald Zambrano for all their help and inspiration throughout the project. We would also like to express our gratitude for all the effort the lab put in providing us quality datasets to work with. Lastly, we would like to thank Dr. Julia Kempe for her support.

References

[1] Sihong Chen, Kai Ma, and Yefeng Zheng. "Med3d: Transfer learning for 3d medical image analysis". In: *arXiv preprint arXiv:1904.00625* (2019).

[2] Xinlei Chen et al. "Improved baselines with momentum contrastive learning". In: *arXiv preprint arXiv:2003.04297* (2020).

[3] Sripad Krishna Devalla et al. "Towards label-free 3D segmentation of optical coherence tomography images of the optic nerve head using deep learning". In: *Biomedical optics express* 11.11 (2020), pp. 6356–6378.

[4] Alexey Dosovitskiy et al. "Discriminative Unsupervised Feature Learning with Convolutional Neural Networks". In: *Advances in Neural Information Processing Systems*. Ed. by Z. Ghahramani et al. Vol. 27. Curran Associates, Inc., 2014. URL: https://proceedings.neurips.cc/paper/2014/file/07563a3fe3bbe7e3ba84431ad9d055af-Paper.pdf.

[5] Shih-Hau Fang et al. "Detection of Pathological Voice Using Cepstrum Vectors: A Deep Learning Approach". In: *Journal of Voice* 33.5 (Sept. 2019), pp. 634–641. DOI: 10.1016/j.jvoice.2018.02.003. URL: https://doi.org/10.1016/j.jvoice.2018.02.003.

[6] Kaiming He et al. "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.

[7] Stefan Maetschke et al. "A feature agnostic approach for glaucoma detection in OCT volumes". In: *PloS one* 14.7 (2019), e0219126.

[8] Riccardo Miotto et al. "Deep learning for healthcare: review, opportunities and challenges". In: *Briefings in Bioinformatics* 19.6 (May 2017), pp. 1236–1246. DOI: 10.1093/bib/bbx044. URL: https://doi.org/10.1093/bib/bbx044.

[9] Mehdi Noroozi and Paolo Favaro. "Unsupervised Learning of Visual Representations by Solving Jigsaw Puzzles". In: *CoRR* abs/1603.09246 (2016). arXiv: 1603.09246. URL: http://arxiv.org/abs/1603.09246.

[10] Abhronil Sengupta et al. "Going deeper in spiking neural networks: VGG and residual architectures". In: *Frontiers in neuroscience* 13 (2019), p. 95.

[11] Mennatullah Siam et al. "Deep semantic segmentation for automated driving: Taxonomy, roadmap and challenges". In: *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, Oct. 2017. DOI: 10.1109/itsc.2017.8317714. URL: https://doi.org/10.1109/itsc.2017.8317714.

[12] Aiham Taleb et al. "3d self-supervised methods for medical imaging". In: *arXiv preprint arXiv:2006.03829* (2020).

[13] Athanasios Voulodimos et al. "Deep Learning for Computer Vision: A Brief Review". In: *Computational Intelligence and Neuroscience* 2018 (2018), pp. 1–13. DOI: 10.1155/2018/7068349. URL: https://doi.org/10.1155/2018/7068349.

[14] Bo Wang et al. "In-vivo effects of intraocular and intracranial pressures on the lamina cribrosa microstructure". In: *PloS one* 12.11 (2017), e0188302.

[15] Zifeng Wu, Chunhua Shen, and Anton Van Den Hengel. "Wider or deeper: Revisiting the resnet model for visual recognition". In: *Pattern Recognition* 90 (2019), pp. 119–133.

[16] Anbang Xu et al. "A New Chatbot for Customer Service on Social Media". In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, May 2017. DOI: 10.1145/3025453.3025496. URL: https://doi.org/10.1145/3025453.3025496.

[17] Tom Young et al. "Recent Trends in Deep Learning Based Natural Language Processing [Review Article]". In: *IEEE Computational Intelligence Magazine* 13.3 (Aug. 2018), pp. 55–75. DOI: 10.1109/mci.2018.2840738. URL: https://doi.org/10.1109/mci.2018.2840738.