# Supplementary Material
# AstroNet: When Astrocyte Meets Artificial Neural Network

Mengqiao Han* Liyuan Pan*,† Xiabi Liu†
Beijing Institute of Technology
hmq@bit.edu.cn

## Abstract

We provide additional implementation details on the design of the AN architecture (Sec. 1), the additional ablation study on the hyperparameters in the AN optimization objective function (Sec. 2), and we visualize the attention regions on the input image with NN under the different structures of AN and the selection of the global connection function (Sec. 3). Furthermore, we validate our performance on the Vision Transformer (ViT) model (Sec. 4) and the segmentation task (Sec. 5) on ImageNet-1k.

## 1. Astrocyte Network Design

The core idea of our design of the Astrocyte Network (short as AN) is that it can effectively integrate the weights of the neural network (short as NN). The output of AN is the probability of each filter or neuron, which is analogous to semantic segmentation, *i.e.*, classifying each element in the feature matrix. Therefore, we naturally design AN as an UNet structure [5] or a fully convolutional network (FCN) structure [4], both of which excel in segmentation tasks. We also report designing the AN structure as a convolutional neural network (CNN) [2] style to demonstrate the effectiveness of our AstroNet architecture. Note that since NN weights are represented by their features, which allows AN to be a relatively small-size network. Specifically, the AN designed based on UNet, FCN and CNN structures are shown in Fig. 1, both of which are composed of a few convolutional layers.

Tab. 1 shows the performance of AstroNet on the testing set with the three AN structure designs. Different structure designs of AN in AstroNet can achieve better test accuracy than the ResNet18 baseline, which proves the effectiveness of our Astrocyte-Neuron model. We also observed that designing AN with a network structure suitable for segmentation tasks will achieve better results. In addition, UNet and FCN structures have another obvious advantage over CNN;

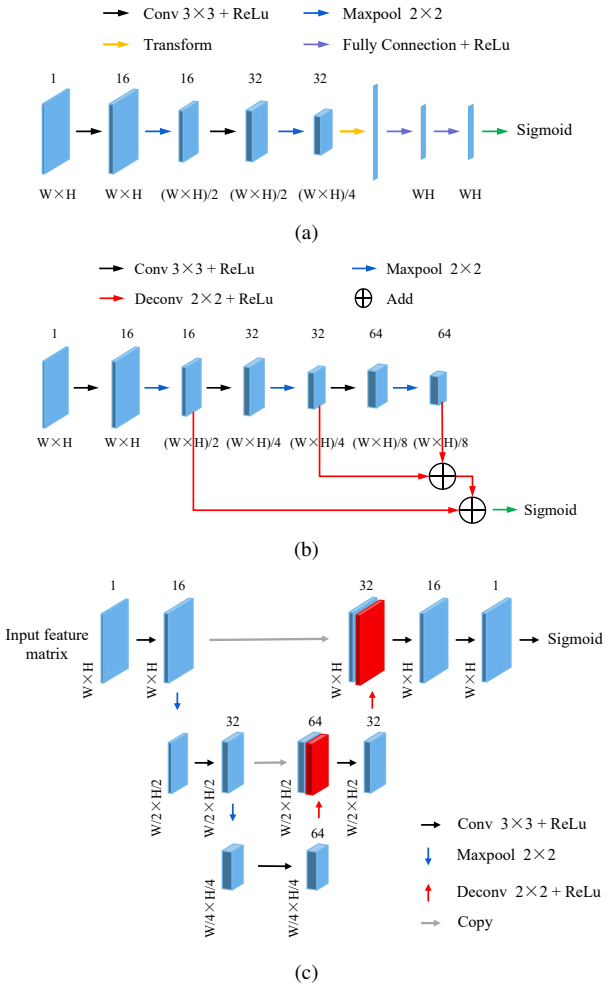



Figure 1. Different AN structure designs. (a) AN is designed as a CNN-based structure. (b) AN is designed as an FCN-based structure. (c) AN is designed as a UNet-based structure. (Best viewed in color on the screen)

their input and output tensors are the same sizes. With UNet and FCN, the basic pattern of the AN structure does not

*Equal contribution, †corresponding authors

Table 1. *Using ResNet18 as the backbone, we measure the test accuracy and parameter quantities of our AstroNet with three AN architectures on the CIFAR10 dataset.*

| AN Structure | Acc (%) | Params (M) |
|---|---|---|
| ResNet18 [7] (Baseline) | 92.80 | 11.17 |
| CNN | 94.21 | 7.9 |
| FCN | $\underline{94.31} \pm 0.17$ | 7.8 |
| UNet | $\mathbf{94.35 \pm 0.14}$ | 7.6 |

need to be adjusted for different input sizes and can be flexibly applied to various network layers.

To illustrate the difference in designing ANs with different structures, we further show the attention regions of NNs (with ResNet18 or DenseNet-BC) on input images. The visualization follows the Grad-CAM [6]. The maximum number of iterations $T$ is set to 6, and all connections establish bidirectional propagation with AN (consistent with our paper). Note, we only display the attention regions after finishing the iteration. Fig. 2 and Fig. 3 show the transfer of attention regions on the input image with NN (ResNet18 or DenseNet-BC) with different structures of AN, and their comparison with the attention regions of standard NNs. Our approach makes the NN pay more attention to the target region on the input image rather than the surrounding environment. In addition, the image in Fig. 2 and Fig. 3 labeled as 'dog', contains two legs and a dog. Our method can pay proper attention to dogs. Designing the AN as the UNet and FCN structures, especially the UNet structure, our AstroNet pays more attention to the target regions in the input image than the CNN structure. The anti-interference ability for environmental regions of UNet-based AN is better as well.

## 2. Effect of Different $\lambda$ Values on Test Accuracy

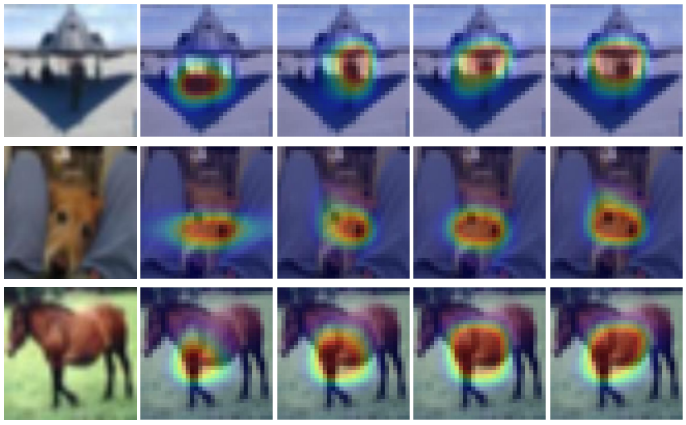
The optimization of AN except for the performance evaluation item, we also introduce a constraint term that reduces the difference between $e^*$ (*i.e.*, minimizing the difference between the output of AN $P^t$), where $e^t = ||P^t - P^{t-1}||_2$, $1 < t \leq T$, $T$ denotes the maximum iteration number. When $t = 1$, $e^1 = ||P^1||_2$. This constraint term regularises our AN to gradually optimize NN steadily. We give the optimization objective of AN through Eq. (10) in our paper, which is expressed as,

$$\begin{aligned} \mathcal{L}_{an} &= E(\mathcal{H}(x, W \odot P^t), y_g) + \lambda ||e^t - e^{t-1}||_2 \,, \\ e^t &= ||P_h^t - P_h^{t-1}||_2 \,, \\ P_h^t &= \mathcal{U}(p^t), \quad P_h^{t-1} = \mathcal{U}(p^{t-1}) \,, \end{aligned} \tag{10}$$

where $||\cdot||_2$ is the $\ell_2$ norm, $\lambda$ is the weight parameter to balance these two terms, and $\mathcal{U}(\cdot)$ denotes the normalization function.

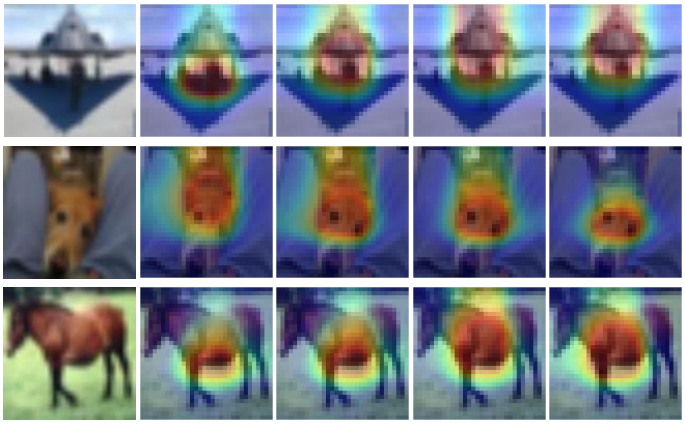
In this section, we conduct an ablation study on different values of $\lambda$. Take NN as ResNet18 and ResNet34 as examples, Fig. 4 shows the $||e^t - e^{t-1}||_2$ with respect to iterations under different $\lambda$. The results indicated that our AstroNet gradually optimizes the network connections against the iteration times from 1 to 10.

Tab. 2 shows the test accuracy with respect to different $\lambda$. Adding a constraint term makes the test accuracy better. Since it makes the working mechanism of AN more



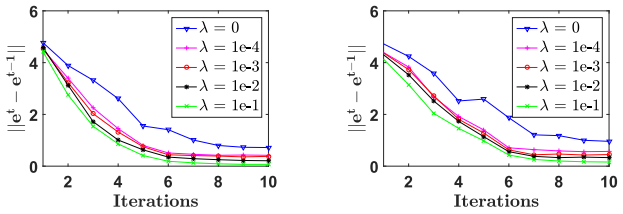
(a) Input (b) Standard (c) CNN (d) FCN (e) Ours

Figure 2. *We illustrate the NN's (ResNet18 on CIFAR10) attention on the input image with different AN structures. From the $1^{st} - 3^{rd}$ rows, the labels are 'Airplane', 'Dog', and 'Horse' (a) Input image. (b) Standard ResNet18 without AN. (c) CNN-based AN structure. (d) FCN-based AN structure. (e) Ours, with a UNet-based AN structure.*



(a) Input (b) Standard (c) CNN (d) FCN (e) Ours

Figure 3. *We illustrate the NN's (DenseNet-BC on CIFAR10) attention on the input image with different AN structures. From the $1^{st} - 3^{rd}$ rows, the labels are 'Airplane', 'Dog', and 'Horse' (a) Input image. (b) Standard DenseNet-BC without AN. (c) CNN-based AN structure. (d) FCN-based AN structure. (e) Ours, with a UNet-based AN structure.*




(a) ResNet18 (b) ResNet34

Figure 4. *Variation of the difference between AN outputs corresponding to different $\lambda$ with iterations.*

Table 2. *Comparison of different $\lambda$ values on ResNet18 and ResNet34 in the CIFAR10 dataset.*

| Method | $\lambda$ | Acc (%) | Params (M) |
|---|---|---|---|
| ResNet18 [7] | - | 92.80 | 11.17 |
| Our-1 | 0 | $94.25 \pm 0.20$ | 7.8 |
| Our-2 | 1e-4 | $\underline{94.33} \pm 0.16$ | 7.7 |
| Our-3 | 1e-3 | $\mathbf{94.35 \pm 0.14}$ | 7.6 |
| Our-4 | 1e-2 | $94.32 \pm 0.13$ | 7.6 |
| Our-5 | 1e-1 | $94.27 \pm 0.11$ | 7.4 |
| ResNet34 [7] | - | 93.56 | 21.28 |
| Our-1 | 0 | $94.42 \pm 0.19$ | 9.9 |
| Our-2 | 1e-4 | $\underline{94.49} \pm 0.17$ | 9.7 |
| Our-3 | 1e-3 | $\mathbf{94.51 \pm 0.13}$ | 9.7 |
| Our-4 | 1e-2 | $94.32 \pm 0.14$ | 9.5 |
| Our-5 | 1e-1 | $94.27 \pm 0.12$ | 9.4 |

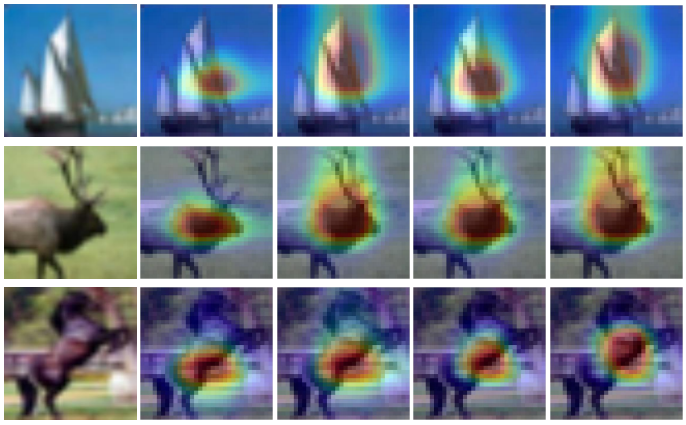
consistent with the temporal regulation mechanism in the Astrocyte-Neuron model, *i.e.*, the regulation provided by astrocytes is gradually decayed.

## 3. Different Global Connection Functions in the Temporal Regulation Mechanism

We discuss the selection of global connection function $\mathcal{G}(\cdot)$ in Tab. 3. Compared to only using $\mathcal{G}_{avg}(\cdot)$ or $\mathcal{G}_{max}(\cdot)$, where $\mathcal{G}_{avg}(\cdot)$ and $\mathcal{G}_{max}(\cdot)$ is the feature matrix of $W$ with the average and maximum connection intensity of the NN. Our settings (mix the $\mathcal{G}_{avg}(\cdot)$ and $\mathcal{G}_{max}(\cdot)$) achieve the best performance. The results indicated that it is more reasonable first to learn the global features of each neuron and then gradually regulate the important connection.

Table 3. *Comparison of global connection function $\mathcal{G}(\cdot)$ in iterations on the CIFAR10 dataset, where NN is set to ResNet18.*

| $\mathcal{G}(\cdot)$ | Acc (%) | Params (M) |
|---|---|---|
| Global average pooling | 94.32 | 7.6 |
| Largest connection | 94.31 | 7.8 |
| Our (Mixture) | **94.35** | 7.6 |



(a) Input (b) Standard (c) $\mathcal{G}_{avg}(\cdot)$ (d) $\mathcal{G}_{max}(\cdot)$ (e) Ours

Figure 5. *We illustrate the NN's (ResNet18 on CIFAR10) attention on the input image with different global connection functions. From the $1^{st}-3^{rd}$ rows, the labels are 'Boat', 'Deer', and 'Horse' (a) Input image. (b) Standard ResNet18 without AN. (c) Global average pooling $\mathcal{G}_{avg}(\cdot)$. (d) Largest connection $\mathcal{G}_{max}(\cdot)$. (e) Ours (Mixture).*

Table 4. *Results of our method on the visual transformer on ImageNet-1k. Our (ViT) denotes the NN in our AstroNet is set to ViT.*

| Architecture | Acc (%) | Params (M) |
|---|---|---|
| ViT [1] | 77.9 | 86 |
| Ours (ViT) | **79.3** | **68** |

Table 5. *Experiments on the segmentation method with different backbones on COCO. ResNet50 and ResNet50† are pre-trained on ImageNet-1k by standard training and our method, respectively.*

| Model | backbone | AP | $AP_S$ | $AP_M$ | $AP_L$ |
|---|---|---|---|---|---|
| BoxInst [8] | ResNet50 | 32.1 | 15.6 | **34.3** | 43.5 |
| Ours | ResNet50† | **33.7** | **15.9** | 34.1 | **44.2** |

We further show the attention of NN (ResNet18) with different global connection functions $\mathcal{G}(\cdot)$ on input images, and their comparison with the attention regions of standard NNs. As shown in Fig. 5, our method (e) observes the edges of the target in more detail than only using $\mathcal{G}_{avg}(\cdot)$ (c). For example, in the $2^{th}$ row labeled as 'Deer', our method notices the antler region. This is related to our additional introduction of $\mathcal{G}_{max}(\cdot)$, which enhances the attention of the NN connection to the target characteristic. Compared to $\mathcal{G}_{max}(\cdot)$ (d), the attention of our method covers more target regions on the input images. However, no matter which global connection function $\mathcal{G}(\cdot)$ is used in our method, the attention of our method to target regions on the input image is better than standard ResNet18 without AN to regulate its connections.

## 4. Experimental Results on the Vision Transformer Model

We apply our method to the vision transformer on ImageNet-1k, to evaluate the effectiveness of our method on different NN architectures. The structure of NN in our AstroNet is ViT [1]. The results are shown in Tab. 4, compare with ViT, our method achieves a relative improvement in accuracy by $1.4\%$. It also reduces the capacity of ViT by $20.9\%$.

## 5. Experimental Results on the Downstream Task

We compare the segmentation task with the SOTA box-supervised method BoxInst [8] on the COCO [3]. BoxInst uses ResNet50 (pre-trained on the ImageNet-1k) as the backbone. We exactly follow the framework of BoxInst, except replace the ResNet50 with our ResNet50† as the backbone. Our ResNet50† is obtained by first searching for AN on the ImageNet-1k and then using AN to guide the training of ResNet50. Tab. 5 reports the performance of our method on COCO. It is observed that the backbone (ResNet50†) pre-trained with our method, compared to BoxInst, achieves AP improvements of $1.6\%$. And our method outperforms BoxInst on most metrics.

## References


[1] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, et al. An image is worth 16x16 words: Transformers for image recognition at scale. In *Int. Conf. Learn. Represent.*, 2021. 3, 4

[2] Jiuxiang Gu, Zhenhua Wang, Jason Kuen, Lianyang Ma, Amir Shahroudy, Bing Shuai, Ting Liu, Xingxing Wang, Gang Wang, Jianfei Cai, et al. Recent advances in convolutional neural networks. *Pattern recognition*, 77:354–377, 2018. 1

[3] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014. 4

[4] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015. 1

[5] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 1

[6] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pages 618–626, 2017. 2

[7] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 2, 3

[8] Z. Tian, C. Shen, X. Wang, and H. Chen. Boxinst: High-performance instance segmentation with box annotations. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 5443–5452, 2021. 3, 4