MA-Net: Rethinking Neural Unit in the Light of Astrocytes

Mengqiao Han, Liyuan Pan†, Xiabi Liu†

Beijing Institute of Technology
{hmq, liyuan.pan, liuxiabi}@bit.edu.cn

Abstract

The artificial neuron (N-N) model-based networks have accomplished extraordinary success for various vision tasks. However, as a simplification of the mammal neuron model, their structure is locked during training, resulting in overfitting and over-parameters. The astrocyte, newly explored by biologists, can adaptively modulate neuronal communication by inserting itself between neurons. The communication, between the astrocyte and neuron, is bidirectional and shows the potential to alleviate issues raised by unidirectional communication in the N-N model. In this paper, we first elaborate on the artificial Multi-Astrocyte-Neuron (MA-N) model, which enriches the functionality of the artificial neuron model. Our MA-N model is formulated at both astrocyte- and neuron-level that mimics the bidirectional communication with temporal and joint mechanisms. Then, we construct the MA-Net network with the MA-N model, whose neural connections can be continuously and adaptively modulated during training. Experiments show that our MA-Net advances new state-of-the-art on multiple tasks while significantly reducing its parameters by connection optimization.

Introduction

Networks constructed with the artificial neuron (N-N) model have achieved notable successes (Han et al. 2022; Cheng et al. 2022; Yang, Pan, and Liu 2023). However, their network structures are fixed once the training starts, *i.e.*, structures cannot be adjusted during training. This brings up two problems: 1) over-fitting (Li et al. 2021; Cao et al. 2022); and 2) over-parameterization (Frankle and Carbin 2019), as the complexity of a network can not apply to all data. Though existences attempt to solve the mentioned issues by constraining and searching neural connections, such still lead to unstable performance and resource consumption (Li et al. 2020; Yang, Liu, and Xu 2023).

The key culprit for the weakness of the N-N model is the simplified unidirectional communication between neurons (McCulloch and Pitts 1943; Rosenblatt 1958; Hai et al. 2023). In previous understanding, pre-neurons release neurotransmitters to stimulate a post-neuron, which is unidirectional propagation. Neurotransmitters are formulated as connections (weights) in a network. Indicating by the new explorations, astrocytes assist the connections between the pre- and post-neuron with bidirectional communication. The bidirectional allows the neurotransmitters, *i.e.*, network connections, to be modulated during training adaptively and iteratively (see Fig. 1)

Existing work starts from modeling a single astrocyte to optimize the network connection, *e.g.*, a neuronal-astrocytic group (Hastings et al. 2023) and AstroNet (Han, Pan, and Liu 2023). Considering the neuron activity varies from each other (Frankle and Carbin 2018), it is difficult to fully mine structural features for all neurons through only one astro-

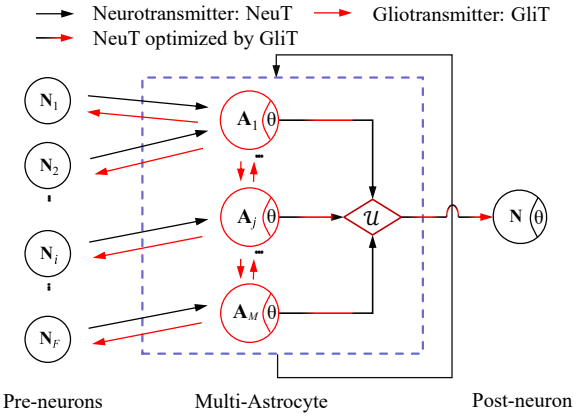


Figure 1: The illustration of our bidirectional communication between neurons and astrocytes. Previous N-N model only let pre-neurons $\{\mathbf{N}_i\}_{i=1}^{F}$ release neurotransmitters to connect to the post-neuron $\mathbf{N}$, where F is the number of pre-neurons. In our model, M astrocytes $\{\mathbf{A}_j\}_{j=1}^{M}$ are involved to jointly modulate the neurotransmitters by releasing gliotransmitters. For example, astrocyte $\mathbf{A}_1$ is stimulated by neurotransmitters released by pre-neurons $\{\mathbf{N}_i\}_{i=1}^{F_1}$ and produce gliotransmitters. The gliotransmitters force these pre-neurons to re-release neurotransmitters (Perea and Araque 2010), achieving the modulation of connections. Moreover, we formulate the modulation iteratively both between 'astrocyte to pre-neurons' and 'astrocyte to astrocyte' (purple box). Our iteratively optimized Multi-Astrocyte-Neuron (MA-N) model, therefore, mimics the functionality of a relatively real neuron model.

† Corresponding author.

cyte. Moreover, their modeled single astrocyte maintains only one state during the iterations, weakening the ability of temporal modulation of the bidirectional communication.

In this paper, we first revisit the bidirectional connection between astrocytes and neurons. Then, pioneering formulates the artificial Multi-Astrocyte-Neuron (MA-N) model at both astrocyte- and neuron-level.

Astrocyte-level: Considering astrocytes are inherently abundant in the nervous system (Perea and Araque 2010), we need to model multi-astrocyte in the MA-N model. Unlike the existing single astrocyte model, we model the multiple sub-astrocytes by two modulation mechanisms. First, our sub-astrocyte varies in iterations to enable its temporal ability. For example, a sub-astrocyte at iteration t is the modulation results of its neighbor sub-astrocytes at iteration $t-1$, as a sub-astrocyte can be updated by historically released gliotransmitters (Fossati, Matteoli, and Menna 2020). Second, we also optimize the interaction between sub-astrocytes at the current iteration jointly, as each sub-astrocyte is not only stimulated by neurotransmitters from its associated partial pre-neurons but also implicitly affected with the rest of the pre-neurons by the interaction between sub-astrocytes. We, therefore, build a temporal and joint modulation mechanism to mimic the interaction among each sub-astrocyte.

Neuron-level: The neuron connections, *i.e.*, neurotransmitters, also collaborate to influence each sub-astrocyte (Fields and Stevens-Graham 2002). On the one hand, each sub-astrocyte is affected by its associated pre-neurons' released neurotransmitter sequences, which reflect the modulation history of the pre-neurons. On the other hand, the association of pre-neurons and sub-astrocyte is implicitly affected by the global connection of all neurons jointly. Hence, we further enhance the temporal and joint modulation mechanism at the neuron-level in our MA-N model.

With the built MA-N model, we construct our MA-Net, where its structure can be optimized during training. Our main contributions are:

- We formulate the artificial Multi-Astrocyte-Neuron (MA-N) model at both astrocyte- and neuron-level with the temporal and joint modulation mechanism.
- We construct our MA-Net which adaptively optimizes the dynamically changing neural connections during training based on the MA-N model.
- We extensively evaluate out MA-Net on multiple tasks, *e.g.*, classification, segmentation, and object detection with public datasets.

Compared to the state-of-the-art (SOTA) methods, our MA-Net improves the accuracy by $0.23\% \sim 2.91\%$ on classification (CIFAR10 and ImageNet-1k). AP improved by $0.7\% \sim 2.0\%$ on segmentation (COCO) and $0.5\% \sim 1.5\%$ on object detection (COCO). The optimized connections with parameters are reduced by $10.72\% \sim 71.25\%$.

Related Work

N-N model-based network fixes the architecture during training. Extensive efforts have been made for network compression and inference, including but not limited to, parameter pruning, knowledge distillation (KD), and dynamic neural network (DyNN). Parameter pruning contains nonstructural pruning disabling the weak connections (Frankle and Carbin 2019) and structural pruning aiming at zeroing out groups of the convolutional filters (Wen et al. 2016), which has a tedious fine-tuning process and lead to an accuracy drop when pursuing high a compression rate (Li, Wang, and Ruiz 2020). KD shows that the student model learns little from some teacher models due to the model capacity gap between them (Zhao et al. 2022). Existing DyNNs are designed in different aspects including sample-wise by adjusting network architectures based on each sample (Mullapudi et al. 2018), spatial-wise by performing adaptive inference with respect to different spatial locations of images (Wang et al. 2019), and temporal-wise dynamism by dynamically allocating less computation to the inputs at unimportant temporal locations (Hansen et al. 2019). Note that DyNN is a data or image (pixel, region, or resolution level)-dependent method whose structure and parameters vary by each sample or different spatial locations of images/features (Han et al. 2021). In contrast, we optimize the structure and parameters of a network by its own connections to fit each dataset.

Neural architecture search methods (NAS) contain network parameters optimization and architecture optimization. Note NAS can be viewed as an 'adaptive' method searching connections from the artificial neuron model-based search space (Xu et al. 2021). The network parameters optimization includes independent optimization (Real et al. 2019) and sharing optimization (Bender et al. 2018). The architecture optimization is to search the network architectures, including: 1) search space defines which architectures can be defined, including global (Zhong et al. 2020) and cell-based (Brock et al. 2017) search spaces; 2) search strategy includes: RL-based methods (Zoph et al. 2018) use the performances of generated architectures as the rewards for training the controller. EA-based methods (Real et al. 2019) search architectures with evolutionary algorithms. Gradient-based methods (Dong and Yang 2019) regard network architecture as a group of learnable parameters; 3) estimation strategy scores the searched architectures (Xu et al. 2021).

Neural communication is based on the Tripartite Synapse concept (Perea, Navarrete, and Araque 2009) between neurons and astrocytes. Astrocytes establish bidirectional communication with neurons, whereby they respond to synaptically-released neurotransmitters, in turn, release gliotransmitters that influence the activity of neuronal synapses (Bonvento and Bolaños 2021). This shows that changes in synaptic activity combine external disturbances and their own state. In our previous work, we focused on modeling a single astrocyte (Han, Pan, and Liu 2023).

In contrast, this paper aims to formulate an MA-N model to better mimic the bidirectional communication between multi-astrocyte and multi-neuron, temporally and jointly.

Proposed Method

In this section, we first introduce the MA-N model. Then, we elaborate on our MA-Net and its training scheme.

Preliminary. The traditional N-N model (McCulloch and Pitts 1943) formulates the unidirectional connection be-

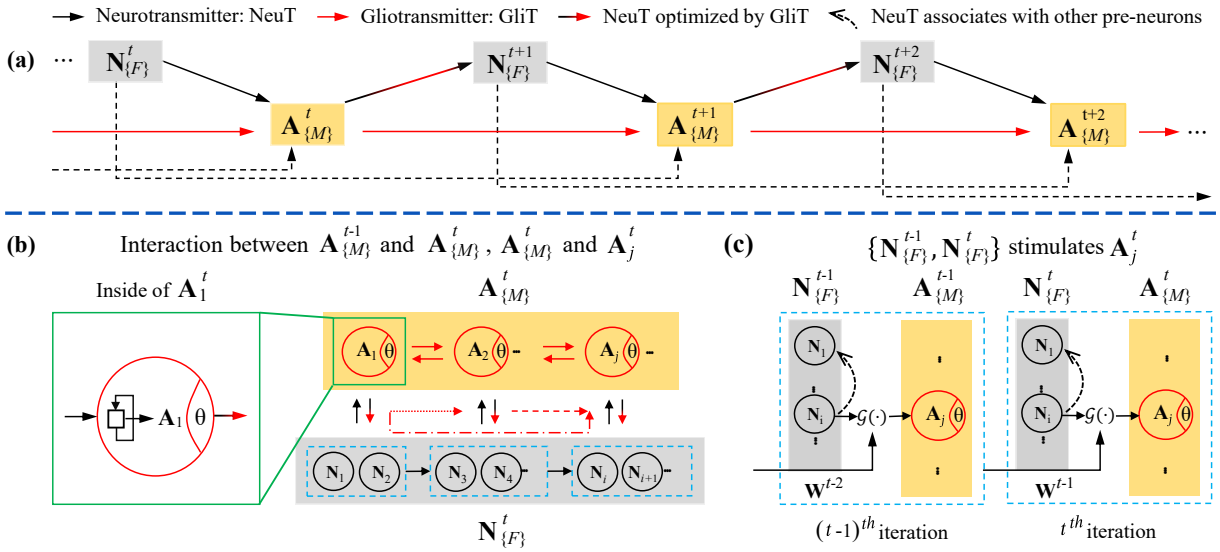


Figure 2: The overall pipeline of our MA-N model. (a) The pipeline of our MA-N model at the astrocyte- and neuron-level with temporal and joint modulation mechanism. (b) At astrocyte-level, a sub-astrocyte $\mathbf{A}^t_j$ at iteration t is modulated by history output of $\{\mathbf{A}^{t-1}_j\}^M_{j=1}$ ($\mathbf{A}^{t-1}_{\{M\}}$ for short) and the interaction among all $\mathbf{A}^t_{\{M\}}$ simultaneously. The green box zoom in as an example of a sub-astrocyte that varies temporally. (c) At the neuron-level, a sub-astrocyte $\mathbf{A}^t_j$ is also stimulated by pre-neurons $\mathbf{N}^t_{\{F_j\}}$ and $\mathbf{N}^{t-1}_{\{F_j\}}$ at iteration t and $t-1$, separately.

tween pre-neurons and post-neuron is formulated as:

$$y = \phi_n\left(\sum_i^F x_i w_i\right) = \phi_n\left(X^\mathsf{T} W\right), \tag{1}$$

where y denotes the output signal of the post-neuron, and $\phi_n()$ is the activation function of the neuron. Let $X = [x_1, \cdots, x_i \cdots, x_F]^\mathsf{T}$ be the data and $W = [w_1, \cdots, w_i \cdots, w_F]^\mathsf{T}$. Here, w_i is the connecting weight from the i^{th} pre-neuron to the post-neuron, and F is the number of pre-neurons.

With the concept of the astrocyte, the neurotransmitters (weights) from pre-neurons are first passed to the astrocyte for modulation. The simplified artificial Astrocyte-Neuron model with only one astrocyte can be formulated as,

$$X^\mathsf{T} W = \sum_{i=1}^F x_i \left(\phi_a\left(f\left(w_i^t\right)\right)\right) = X^\mathsf{T}\left(\phi_a\left(f\left(W^t\right)\right)\right), \tag{2}$$

where $f()$ and $\phi_a()$ denote the modulation and activation function of the astrocyte separately. Here. $W^t = [w_1^t, \cdots, w_i^t, \cdots, w_F^t]^\mathsf{T}$ and $w_i^t = f(w^{t-1})$ is the connecting weight of i^{th} pre-neuron at iteration time $t \in [1, T]$. Previous work (Han, Pan, and Liu 2023) used one state $f()$ during all iterations for all pre-neurons.

Different from the simplified version of the artificial Astrocyte-Neuron model in Eq. (2), we build our MA-N model among both astrocyte- and neuron-level (see Fig. 2a).

Artificial Multi-Astrocyte-Neuron Model

Given M numbers of astrocytes, each one corresponds to a sub-set of F numbers of pre-neurons. For example, the j^{th} astrocyte $\mathbf{A}_j$ establishes the connections with F_j numbers of pre-neurons. Formally,

$$X^\mathsf{T} W = X^\mathsf{T}\left(\phi_a\left(f_j\left(\mathbf{e}_j \odot W^t\right)\right)\right)^M_{j=1}, \tag{3}$$

where $\odot$ is the element-wise product, $f_j()$ denotes the modulation function of the j^{th} astrocyte, and $\mathbf{e}_j \in \mathbb{R}^{F\times 1}$ is a binary vector with value one in the established connection between pre-neurons and astrocytes. Note that $\sum_{j=1}^M \mathbf{e}_j = \mathbb{1}$, $|\mathbf{e}_j| = F_j$. Here, $\mathbf{e}_j$ is the mask for selecting pre-neurons' connecting weights from W^t that participate in bidirectional connections with the j^{th} astrocyte.We further formulate the MA-N model at both the astrocyte- and neuron-level.

Astrocyte-level: The interaction between sub-astrocytes involves two aspects, *i.e.*, temporally and jointly. First, a sub-astrocyte $\mathbf{A}^t_j$ at t^{th} iteration is stimulated by historically gliotransmitters released from $\mathbf{A}^{t-1}_{\{M\}}$ at $(t-1)^{th}$ iterations, and then resulting the modulation of $\mathbf{A}^t_j$. Therefore, it enables us to improve the previous modulation function $f_j()$ to $f^t_j()$ with temporal ability. Second, sub-astrocytes are connected with each other, *i.e.*, a sub-astrocyte is stimulated by transmitters released from neighboring sub-astrocytes. Specifically, a sub-astrocyte $\mathbf{A}^t_j$ is optimized among interactions of all M numbers of astrocytes $\mathbf{A}^t_{\{M-1\}}$ (see Fig. 2b). Therefore, we build a joint function $\mathcal{U}()$ between astrocytes. Based on Eq. (3), we have

$$X^\mathsf{T} W = X^\mathsf{T}\mathcal{U}\left(\phi_a\left(f^t_j\left(\mathbf{e}_j \odot W^t\right)\right)\right)^M_{j=1}, \tag{4}$$

where $f^t_j()$ denotes the temporal modulation function of the j^{th} astrocyte at the iteration time t.

Neuron-level: As each sub-astrocyte modulates a sub-set of pre-neurons, the performance of the sub-astrocyte is infected with neurotransmitters released by associated pre-neurons, *e.g.*, $\mathbf{A}_j^t$ with $\mathbf{N}_{\{F_j\}}^t$ (see Fig. 2c). To modulate the connection of sub-set pre-neurons $\mathbf{N}_{\{F_j\}}^t$ jointly, we build an adaptive function $\mathcal{G}()$. Note, the activity of an astrocyte combines external disturbances and their own state history (Kofuji and Araque 2021). Hence, the $\mathcal{G}()$ not only implicitly connected with $\mathbf{N}_{\{F_j\}}^t$ but also implicitly considers the connection with the rest $F-F_j$ pre-neurons. In addition, historical connections are included, *i.e.*, W^{t-1} of $\mathbf{N}_{\{F\}}^{t-1}$. With the above concerns, we enable the astrocyte to encode both local and global information of the connections with pre-neurons by

$$X^{\mathsf{T}}W = X^{\mathsf{T}}\mathcal{U}\bigg(\phi_a\Big(f_j^t\big(\mathbf{e}_j \odot \mathcal{G}(W^t, W^{t-1})\big)\Big)\bigg)_{j=1}^{M}, \quad (5)$$

where the adaptive function $\mathcal{G}(W^t, W^{t-1})$ is a fusion, *e.g.*, average, of $g_1(w_{\{F_j\}}^t, w_{\{F_j\}}^{t-1})$ and $g_2(W^t, W^{t-1})$ that jointly enhance the temporal ability of each astrocyte. We name the Eq. (5) as our MA-N model.

Framework of MA-Net

Network Architecture: Based on our MA-N model that can modulate network connections adaptively, we build our MA-Net. Given a set of input $\mathbf{X}$, and rearranging Eq. (5) to Eq. (1), yields

$$\mathbf{Y} = \mathcal{N}\Big(\mathbf{X},\ \mathcal{A}\big(\mathbf{W}, \Phi\big)\Big)\ . \quad (6)$$

In our framework, the artificial neurons with function $\mathcal{N}()$ aim to perform various downstream tasks, *i.e.*, $\mathcal{N}()$ is our task network (TNet). Here, TNet is an N-N model-based network with connecting weights $\mathbf{W}$. The $\mathbf{W}$ is optimized by our astrocytes with the function $\mathcal{A}()$, *i.e.*, the modulation network (MNet). The connecting weights of astrocytes $\Phi = \{\phi_{\mathcal{G}}, \phi_f, \phi_{\mathcal{U}}\}$ are implied in $\mathcal{G}()$, $\{f_j^t()\}_{j,t=1}^{M,T}$ and $\mathcal{U}()$.

The $\mathcal{G}()$ first compresses the TNet connections $\mathbf{W}$ before correlating. This is due to the fact that $\mathbf{W}$ can contain large amounts of parameters, *e.g.*, a 101-layer ResNet has around 44.5 million parameters. We express the $\mathcal{G}()$ through an adaptive global compression, *i.e.*, $g_1()$ and $g_2()$ in $\mathcal{G}()$ are realized by the attention module and ViT module, respectively. The attention module (Hu, Shen, and Sun 2018) is applied to each astrocyte-input connection. First, the squeeze operation is used to compress the local receptive field into channel descriptor (feature) through global average pooling. Then, using the fusion operation to fuse these descriptors through two fully connected layers, *i.e.*, a global compressed feature with respect to the input of each astrocyte is obtained. The ViT module (Dosovitskiy et al. 2021) acts on a global compressed input between astrocytes. ViT, by treating the input features of each astrocyte as a patch, represents their relationship in the whole $\mathcal{N}()$ by imposing location information. This enables $\mathcal{G}()$ to integrate the corresponding local and global connectivity features. Note that our ViT can be designed as a lightweight model since the weights have been compressed by the attention module.

For $f_j^t()$ with each sub-astrocyte at an iteration, to enhance its temporal ability, we design it based on LSTM with a recursive structure, *e.g.*, $f_j^t() = [\text{LSTM}^k]_{k=1}^K$. The input of each LSTM layer in sub-astrocyte integrates the output of all previous LSTM layers, *i.e.*, $f_j^t(w^t, k) = \sum_{k=1}^{k-1} f_j^t(w^t, k)$, to enhance astrocytes response to every optimization history.

For $\mathcal{U}()$ that controls the whole astrocyte set, we achieve its functional ability with a chain optimization strategy. As shown in Fig. 2b, in each iteration, we train from $\mathbf{A}_1$ to $\mathbf{A}_M$ until the optimization of the MNet is completed. Therefore, the realization of $\mathcal{U}()$ involves the optimization of $\{f_j^t()\}_{j,t=1}^{M,T}$. The chain optimization brings the temporal correlation between connections by modulating connections to have a temporal impact on other connections, thereby expressing the joint modulation between astrocytes.

Training Scheme: The optimization of our MA-Net is in an alternate manner. Following (Fossati, Matteoli, and Menna 2020), neurons preferentially develop in the neurons-astrocytes system. Therefore, we first optimize TNet such that the TNet weights include task-specific information. The optimal parameter $\mathbf{W}$ is calculated by minimizing the following function:

$$\mathbf{W}^* = \arg\min_{\mathbf{W}} \mathcal{L}_{\text{TN}}(\mathcal{N}(\mathbf{X}, \mathbf{W}), \mathbf{Y}_{\text{gt}})\ , \quad (7)$$

where $\mathcal{L}_{\text{TN}}()$ is the task specific loss function and $\mathbf{Y}_{\text{gt}}$ is the ground-truth label.

Then, with the initialized $\mathbf{W}^*$, we optimized the MNet with a data term $\mathcal{L}_{\text{data}}()$ and regularization term $\mathcal{L}_{\text{reg}}()$. The data term forces the MNet to modulate the TNet for high performance and is defined as

$$\mathcal{L}_{\text{data}}(\Phi) = \mathcal{L}_{\text{TN}}\bigg(\mathcal{N}\big(\mathbf{X},\ \mathcal{A}(\mathbf{W}^*, \Phi)\big), \mathbf{Y}_{\text{gt}}\bigg)\ . \quad (8)$$

The regularization term, inspired by (Zhang et al. 2022), encourages astrocytes to learn similar representations for different features by MNet, *e.g.*, z_j and z_{M+j}, from the same connecting weight of TNet, $\mathbf{W}$.

$$\mathcal{L}_{\text{reg}}(\Phi) = -\sum_{j=1}^{M} \log \frac{\exp(z_j \cdot z_{M+j})/\tau}{\sum_{l=1}^{2M} \mathbb{1}_{[l \neq j]} \exp(z_j \cdot z_l)/\tau}\ , \quad (9)$$

where $\mathbb{1} \in \{0, 1\}$ is an indicator function evaluating to 1 if $l \neq q$ and τ is a temperature. Taking Eq. (8) and (9), the final learning objective of MNet is formulated as

$$\begin{aligned} \Phi^* &= \arg\min_{\Phi}\big(\mathcal{L}_{\text{data}}(\Phi) + \lambda\mathcal{L}_{\text{reg}}(\Phi)\big) \\ &= \arg\min_{\Phi}\big(\mathcal{L}_{\text{MN}}(\Phi)\big)\ , \end{aligned} \quad (10)$$

where λ is to balance the two items. Compared with AstroNet (Han, Pan, and Liu 2023) with carefully tuned initialization, our alternate optimization strategy (se Algorithm 1) is more effective. After MA-Net training, we remove the TNet connections whose probabilities are less than the pruning threshold δ and obtain the final TNet weights.

With the obtained final TNet structure, we then perform on multiple public datasets and achieve high accuracy.

Algorithm 1: The MA-Net Optimization Algorithm

Input: TNet and MNet parameters: $\mathbf{W}$ and Φ, ground-truth label: $\mathbf{Y}_{\mathrm{gt}}$, max iterations: R, and number of rounds to optimize the TNet without MNet: R_{TN}.
Output: TNet output: $\mathcal{N}$, MA-Net output: $\mathbf{Y}$
for $\mathbf{r} = 0$ **to** R:
  if $\mathbf{r} \leq R_{\mathrm{TN}}$:
    Optimize $\mathcal{L}_{\mathrm{TN}}(\mathcal{N}, \mathbf{Y}_{\mathrm{gt}}) \rightarrow$ Update $\mathbf{W}$
  else:
    if $\mathbf{r}\%2 == 0$:
      Optimize $\mathcal{L}_{\mathrm{MN}}(\mathbf{Y}, \mathbf{Y}_{\mathrm{gt}}) \rightarrow$ Update Φ
    else:
      Optimize $\mathcal{L}_{\mathrm{TN}}(\mathbf{Y}, \mathbf{Y}_{\mathrm{gt}}) \rightarrow$ Update $\mathbf{W}$
end for

Experiments

Implementation Details. Our MA-Net is implemented in PyTorch and trained via SGD optimizer with a learning rate of $1e$-1. We set the number of astrocytes as $M = 4$, each astrocyte contains $K = 4$ LSTMs inside, $\lambda = 1e$-1 in Eq. (10), and pruning threshold $\delta = 1e$-3. All report figures are the average of five times repeated experiments. Our code will be published on GitHub for reproducible research.

Image Classification. We validate our MA-Net on two classification datasets, CIFAR10 (Krizhevsky, Hinton et al. 2009) and ImageNet-1k (Deng et al. 2009), with commonly used convolutional neural networks as the TNet including ResNet (RNT) (He et al. 2016), Wide ResNet (WRN) (Zagoruyko and Komodakis 2016), DenseNet-BC (DNT-BC) (Huang et al. 2017), and vision transformers including ViT (Dosovitskiy et al. 2021) and Swin (Liu et al. 2021). We further compare our MA-Net with SOTA NAS methods (Wang et al. 2021; Xiao et al. 2022), DyNN (Yu et al. 2023) method, and N-N model-based CDS (Zhang et al. 2022) with contrastive learning to verify our adaptive ability.

Segmentation. For segmentation, our method is evaluated on COCO (Lin et al. 2014). The comparison with the box-supervised and fully-supervised methods by updating their backbones that are trained under our proposed framework.

Object Detection. For object detection, our method is also evaluated on COCO with RCNN (Ren et al. 2015) and DETR (Dai et al. 2021).

Experimental Results

Image Classification. The experimental results of our MA-Net on CIFAR10 and ImageNet-1k are shown in Tab. 1 and Tab. 2. For CIFAR10, our MA-Net achieves SOTA accuracy on RNT18, RNT50, and WRN50 and outperforms CDS (Zhang et al. 2022) and AstroNet (Han, Pan, and Liu 2023) by 0.35% and 0.33% in accuracy on average. Notably, our method reduces the capacity of RNT18, RNT50 and WRN50 by 37.87%, 71.25% and 43.56%, respectively. For ImageNet-1k, on average, our MA-Net outperforms the CDS and AstroNet by 0.57% and 0.55% in accuracy, respectively. Our method reduces the average capacity of the model by 58.51%. By optimizing the connections in the same TNet, MA-Net can find a larger network structure for complex datasets, indicating that MA-Net can adaptively search more connections to represent complex features.

Architecture	Acc (%)	Params (M)
RNT18	94.96	11.17
CDS	96.49	11.17
AstroNet	<u>96.52</u>	<u>7.31</u>
Ours (RNT18)	**96.81**	**6.94**
RNT50	95.07	25.60
CDS	<u>96.78</u>	25.60
AstroNet	96.76	<u>7.64</u>
Ours (RNT50)	**97.10**	**7.36**
WRN50	95.01	68.90
CDS	96.88	68.90
AstroNet	<u>96.90</u>	<u>42.55</u>
Ours (WRN50)	**97.25**	**38.89**

Table 1: Experiments on classification with different methods on CIFAR10. With the same settings, compare networks constructed by our MA-N model and other models. Our (RNT) and Our (WRN) denote that the TNet in MA-Net is set to RNT or WRN. We highlight the best and the second-best numbers in bold and underlined.

Architecture	Acc (%)	Params (M)
RNT18	69.21	11.68
CDS	72.85	11.68
AstroNet	<u>72.82</u>	<u>8.05</u>
Ours (RNT18)	**73.27**	**7.52**
RNT50	75.30	26.11
CDS	78.25	26.11
AstroNet	<u>78.31</u>	<u>8.70</u>
Ours (RNT50)	**78.96**	**8.16**

Table 2: Experiments on classification with different methods on ImageNet-1k. With the same settings, compare networks constructed by our MA-N model and other models.

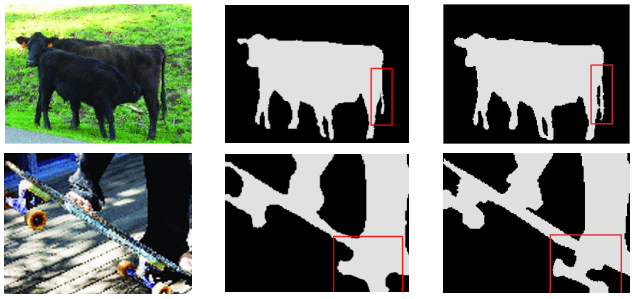
Segmentation. We compare with the box-supervised method BoxInst and the fully-supervised method CondInst on COCO. Note, the backbone, *i.e.*, RNT50, of the standard segmentation method is pre-trained on ImageNet-1k. Keeping the same as other settings in BoxInst (Tian et al. 2021) and CondInst (Tian, Shen, and Chen 2020), we only replace RNT50 with RNT50-CDS, RNT50-Ast and RNT50†, which are pre-trained by CDS, AstroNet and our method, respectively. Tab. 3 reports the comparison of our method with baselines on COCO dataset.

It is observed that the backbone (RNT50†) pre-trained with our method, compared to the second-best, achieves AP improvements of 2.0% and 0.7%, respectively. Furthermore, our RNT50† can reduce the parameters of the segmentation model by 17.95M (see Tab. 2 details). Fig. 3 shows our qualitative segmentation comparison against previous backbones, confirming our benefit. We predict a sharper and completed boundary compared to the AstroNet method, especially on the oxtails and skateboard.

Object Detection. Tab. 4 shows our object detection performance on COCO. We use the same RNT50-based backbone as those in the segmentation task and then fine-tune them

Model	backbone	AP	AP_S	AP_M
BoxInst	RNT50	32.1	15.6	34.3
CDS	RNT50-CDS	32.5	15.9	34.6
AstroNet	RNT50-Ast	32.7	15.7	34.1
Ours	RNT50†	**34.7**	**16.4**	**34.8**
CondIns	RNT50	39.1	21.5	41.7
CDS	RNT50-CDS	39.8	21.6	42.2
AstroNet	RNT50-Ast	40.0	21.8	41.8
Ours	RNT50†	**40.7**	**22.7**	**42.8**

Table 3: Experiments on segmentation with different backbones on COCO. RNT50, RNT50-CDS, RNT50-Ast and RNT50† are pre-trained on ImageNet-1k by standard training, CDS, AstroNet and Ours.



(a) Input (b) AstroNet (c) Ours

Figure 3: We display the segmentation results on images (a) using different backbones, which are obtained by the AstroNet method (b) and our method (c). Our segmentation boundary is accurate, *e.g.*, areas in the red rectangular.

with the same settings for the object detection task. It is observed that by using the backbone (RNT50†) pre-trained in our method, we achieve 1.5% and 0.5% AP improvements compared to the second-best, *i.e.*, AstroNet (RNT50-Ast) in RCNN (Ren et al. 2015) and CDS (Dai et al. 2021) in DETR, respectively.

Discusssions

In this section, we verify the performance of our MA-Net by setting TNet to the vision transformer architecture. Then, we compare the adaptive modulation of neural connections with SOTA dynamic networks. Finally, we compare our transferability with the N-N model-based NAS methods.

With vision transformer. We apply our method to the vision transformer, on ImageNet-1k, to evaluate the effectiveness of our method. The results are shown in Tab. 5. Compared with ViT (Dosovitskiy et al. 2021) and Swin (Liu et al. 2021), by following their settings for training, our method achieves a relative improvement in accuracy by 1.57% and 1.16%, respectively. It also reduces the capacity of ViT and Swin by 19.33% and 10.72%. For the popular vision transformer, our method shows the potential of our MA-Net to alleviate the high parameter problem of transformers.

Comparison with DyNN method. We evaluate the performance of our connection-based adaptive modulation method with the sample (feature)-based adaptive inference DyNN method (Yu et al. 2023). Compared with SOTA Boost-DyNN (76.05%), our MA-Net (78.96%) still achieves competitive performance with more network parameters reduced.

Model	backbone	AP	AP_S	AP_M
RCNN	RNT50	37.4	21.2	41.0
CDS	RNT50-CDS	38.3	21.6	42.0
AstroNet	RNT50-Ast	38.0	21.3	42.2
Ours	RNT50†	**39.8**	**21.9**	**42.5**
DETR	RNT50	42.9	24.6	44.9
CDS	RNT50-CDS	43.2	25.1	45.4
AstroNet	RNT50-Ast	43.5	25.0	45.5
Ours	RNT50†	**44.0**	**25.8**	**45.8**

Table 4: Experiments on object detection with different backbones on COCO. RNT50, RNT50-CDS, RNT50-Ast and RNT50† are pre-trained on ImageNet-1k by standard training, CDS, AstroNet and Ours.

Architecture	Acc (%)	Params (M)
ViT	77.90	86.20
Ours (ViT)	**79.47**	**69.54**
Swin	84.50	84.70
Ours (Swin)	**85.66**	**75.62**

Table 5: Results of our method on the visual transformer on ImageNet-1k. Our (ViT) or Our (Swin) denotes that the TNet in MA-Net is set to ViT or Swin, respectively.

Architecture	Acc (%)	Params (M)
RNT50	75.30	25.60
Boost-DyNN	76.05	11.38
Ours (RNT50)	**78.96**	**8.16**

Table 6: Compared our method with the SOTA DyNN method on ImageNet-1k. Take RNT50 whose performance is close to that of the DyNN method as a reference.

Comparison with NAS methods. Tab. 7 reports the performance results of our MA-Net and SOTA NAS methods on CIFAR10. Compared with NAS-LID (He et al. 2023) and HOTNAS Yang, Liu, and Xu (2023), our method improves the accuracy by 0.25% and 0.23% with fewer computational resources and time costs. The comparison results on ImageNet-1k are demonstrated in Tab. 8. We trained the best-found architectures (both our MA-Net and NAS) on CIFAR10 to evaluate their transferability on ImageNet-1k. Compared to the second-best method HOTNAS, our method leads to an improvement in accuracy by 0.96%, which verifies the transferability of our method.

Comparison with the plug-and-play module. we show comparison results with the plug-and-play module on ImageNet-1k, like non-local and channel attention. As shown in Tab. 9, we achieve competitive accuracy compared to these plug-and-play modules. In comparison to the non-local attention method (Chen et al. 2023), our method yields an accuracy improvement of 0.87%. Similarly, when compared with the channel attention method (Jin et al. 2022), we observe a more substantial enhancement of 1.23% in ac-

Architecture	Acc (%)	Params (M)	Time Cost (GPU days)
DNT-BC	96.54	25.60	-
NAS-LID	97.52	6.90	2.30
HOTNAS	97.54	4.70	3.40
Ours (DNT-BC)	**97.77**	7.52	**0.65**

Table 7: Comparison with SOTA NAS methods. MA-Net and NAS architectures are searched and evaluated on CIFAR10. Time cost only denotes the search time by our method and NAS methods.

Architecture	Acc (%)	Params (M)	Time Cost (GPU days)
RNT50	75.30	26.11	-
NAS-LID	77.10	6.90	2.30
HOTNAS	77.30	4.70	3.40
Ours (RNT50)	**78.26**	6.94	**0.62**

Table 8: Comparison with NAS methods for transferability. MA-Net and NAS architectures are searched on CIFAR10 and then evaluated on ImageNet-1k.

Architecture	Params (M)	Acc (%)
ResNet50	26.11	75.30
ResNet50 + non-local	27.22	78.09
ResNet50 + channel	28.19	77.73
Ours (ResNet50)	**8.16**	**78.96**

Table 9: Comparison with different plug-and-play modules on ImageNet-1k.

curacy. Notably, our method markedly reduces the number of parameters.

Mores. For the varying parameters, 1) the capacity of our MA-Net is related to the size of the chosen TNet. To this end, we show the results with different numbers of parameters (Tiny, Organ, and Large) based on DNT-BC in Tab. 10. Our method still achieves competitive performance on the tiny DNT-BC, and the capacity is significantly reduced; 2) we focus on finding the optimal architecture. The pruning threshold δ can be increased to further reduce parameters.

Ablation Studies

This section discusses the rationality of our method, where the TNet in MA-Net is set to RNT18 in all experiments and evaluated on CIFAR10.

The number of astrocytes in MNet. We investigate the MA-Net performance with respect to the numbers of sub-astrocyte in the MNet. Tab. 11 shows a positive correlation between the number and performance of astrocytes, *i.e.*, increasing astrocyte numbers can achieve higher accuracy. Moreover, reducing astrocyte input, the associated sub-neuron set, allows us to design sub-astrocyte capacity to be smaller. Considering the time cost, we set the number of astrocytes to 4.

The architecture of $\mathcal{U}()$. The temporal-based joint function $\mathcal{U}()$ allows each sub-astrocyte to be modulated by the other astrocytes and their historical output. We set up progressive ablation experiments with three cases to verify the effectiveness of our chain optimization strategy. Case 1: we use independent optimization for multiple astrocytes in MNet, *i.e.*, each astrocyte independently modulates the connections of a sub-set pre-neurons. Case 2: we allow independent modulation of astrocytes to be based on their output at different iterations. Case 3: we use chain optimization for multiple astrocytes to establish interactions between astrocytes. Ours: we allow this interaction to be based on the output of astrocytes at different times. Tab. 12 shows the performance with three cases. Note, all cases achieve better accuracy than the RNT18 baseline, and our designed temporal-based chain optimization achieves the best performance.

Pattern	Architecture	Acc (%)	Params (M)
Tiny-DNT-BC	64-128-128-256	97.55	**7.25**
Orgn-DNT-BC	64-128-256-512	97.77	7.52
Large-DNT-BC	64-256-256-512	**97.86**	7.94

Table 10: Comparison of different sizes of parameters of TNet in MA-Net on CIFAR10. We set the TNet as DNT-BC with different sizes. Taking '64-128-256-512' as an example, it denotes the output of the four dense blocks.

Astrocyte Numbers	AstroNet	2	4	6
Acc (%)	96.52	96.70	**96.81**	**96.81**
Paramrs (M)	2.88	2.52	5.04	7.56
Time Cost (GPU days)	0.07	0.10	0.17	0.26

Table 11: Comparing the accuracy of MA-Net with respect to different numbers of astrocytes on CIFAR10. AstroNet sets a single astrocyte to UNet. The time cost of AstroNet is the result of iteratively modulating the connections 6 times. Our MA-Net, therefore, sets 'sub-astrocyte with 6 LSTMs.

Manners	Optimization for Multi-Astrocyte	Acc (%)	Params (M)
RNT18	-	94.96	11.17
Case 1	Independent	96.54	7.33
Case 2	Independent	96.57	7.29
Case 3	Chain	96.79	6.97
Ours (RNT18)	Chain	**96.81**	**6.94**

Table 12: Experiments on classification with different temporal-based joint functions $\mathcal{U}()$ on CIFAR10.

Conclusions

In this paper, we propose the MA-N model by studying bidirectional connections between multiple astrocytes and neurons, from which we construct MA-Net. By analyzing the bidirectional connection mechanism between astrocytes and neurons, we formulate the connective enhancement mechanism to improve its bidirectional ability, and the joint modulation mechanism between astrocytes, allowing our MA-Net to modulate its connections adaptively. Experiments on multiple tasks and datasets demonstrate that our MA-Net achieves SOTA accuracy and significantly reduces the network's parameters with adaptive optimization.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (82171965, 62302045) and the Beijing Institute of Technology Research Fund Program for Young Scholars.

References

Bender, G.; Kindermans, P.-J.; Zoph, B.; Vasudevan, V.; and Le, Q. 2018. Understanding and simplifying one-shot architecture search. In *International conference on machine learning*, 550–559. PMLR.

Bonvento, G.; and Bolaños, J. P. 2021. Astrocyte-neuron metabolic cooperation shapes brain activity. *Cell metabolism*, 33(8): 1546–1564.

Brock, A.; Lim, T.; Ritchie, J. M.; and Weston, N. 2017. Smash: one-shot model architecture search through hypernetworks. *arXiv preprint arXiv:1708.05344*.

Cao, Y.; Chen, Z.; Belkin, M.; and Gu, Q. 2022. Benign overfitting in two-layer convolutional neural networks. *Advances in neural information processing systems*, 35: 25237–25250.

Chen, F.; Datta, G.; Kundu, S.; and Beerel, P. A. 2023. Self-Attentive Pooling for Efficient Deep Learning. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 3974–3983.

Cheng, B.; Misra, I.; Schwing, A. G.; Kirillov, A.; and Girdhar, R. 2022. Masked-attention mask transformer for universal image segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 1290–1299.

Dai, X.; Chen, Y.; Yang, J.; Zhang, P.; Yuan, L.; and Zhang, L. 2021. Dynamic detr: End-to-end object detection with dynamic attention. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2988–2997.

Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; and Fei-Fei, L. 2009. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, 248–255. Ieee.

Dong, X.; and Yang, Y. 2019. Searching for a robust neural architecture in four gpu hours. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1761–1770.

Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. 2021. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In *International Conference on Learning Representations*.

Fields, R. D.; and Stevens-Graham, B. 2002. New insights into neuron-glia communication. *Science*, 298(5593): 556–562.

Fossati, G.; Matteoli, M.; and Menna, E. 2020. Astrocytic factors controlling synaptogenesis: a team play. *Cells*, 9(10): 2173.

Frankle, J.; and Carbin, M. 2018. The Lottery Ticket Hypothesis: Finding Sparse, Trainable Neural Networks. In *International Conference on Learning Representations*.

Frankle, J.; and Carbin, M. 2019. The Lottery Ticket Hypothesis: Finding Sparse, Trainable Neural Networks. In *International Conference on Learning Representations*.

Hai, Z.; Pan, L.; Liu, X.; Liu, Z.; and Yunita, M. 2023. L2T-DLN: Learning to Teach with Dynamic Loss Network. In *Thirty-seventh Conference on Neural Information Processing Systems*.

Han, K.; Wang, Y.; Chen, H.; Chen, X.; Guo, J.; Liu, Z.; Tang, Y.; Xiao, A.; Xu, C.; Xu, Y.; et al. 2022. A survey on vision transformer. *IEEE transactions on pattern analysis and machine intelligence*, 45(1): 87–110.

Han, M.; Pan, L.; and Liu, X. 2023. AstroNet: When Astrocyte Meets Artificial Neural Network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 20258–20268.

Han, Y.; Huang, G.; Song, S.; Yang, L.; Wang, H.; and Wang, Y. 2021. Dynamic neural networks: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11): 7436–7456.

Hansen, C.; Hansen, C.; Alstrup, S.; Simonsen, J. G.; and Lioma, C. 2019. Neural Speed Reading with Structural-Jump-LSTM. In *International Conference on Learning Representations*.

Hastings, N.; Yu, Y.-L.; Huang, B.; Middya, S.; Inaoka, M.; Erkamp, N. A.; Mason, R. J.; Carnicer-Lombarte, A.; Rahman, S.; Knowles, T. P.; et al. 2023. Electrophysiological In Vitro Study of Long-Range Signal Transmission by Astrocytic Networks. *Advanced Science*, 2301756.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

He, X.; Yao, J.; Wang, Y.; Tang, Z.; Cheung, K. C.; See, S.; Han, B.; and Chu, X. 2023. NAS-LID: Efficient Neural Architecture Search with Local Intrinsic Dimension. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 7839–7847.

Hu, J.; Shen, L.; and Sun, G. 2018. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 7132–7141.

Huang, G.; Liu, Z.; Van Der Maaten, L.; and Weinberger, K. Q. 2017. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4700–4708.

Jin, X.; Xie, Y.; Wei, X.-S.; Zhao, B.-R.; Chen, Z.-M.; and Tan, X. 2022. Delving deep into spatial pooling for squeeze-and-excitation networks. *Pattern Recognition*, 121: 108159.

Kofuji, P.; and Araque, A. 2021. G-protein-coupled receptors in astrocyte-neuron communication. *Neuroscience*, 456: 71–84.

Krizhevsky, A.; Hinton, G.; et al. 2009. Learning multiple layers of features from tiny images.

Li, X.; Wang, Y.; and Ruiz, R. 2020. A survey on sparse learning models for feature selection. *IEEE transactions on cybernetics*.

Li, Y.; Gu, S.; Mayer, C.; Gool, L. V.; and Timofte, R. 2020. Group sparsity: The hinge between filter pruning and decomposition for network compression. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 8018–8027.

Li, Y.; Lin, S.; Liu, J.; Ye, Q.; Wang, M.; Chao, F.; Yang, F.; Ma, J.; Tian, Q.; and Ji, R. 2021. Towards compact cnns via collaborative compression. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 6438–6447.

Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; and Zitnick, C. L. 2014. Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, 740–755. Springer.

Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; and Guo, B. 2021. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, 10012–10022.

McCulloch, W. S.; and Pitts, W. 1943. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4): 115–133.

Mullapudi, R. T.; Mark, W. R.; Shazeer, N.; and Fatahalian, K. 2018. Hydranets: Specialized dynamic architectures for efficient inference. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 8080–8089.

Perea, G.; and Araque, A. 2010. GLIA modulates synaptic transmission. *Brain research reviews*, 63(1-2): 93–102.

Perea, G.; Navarrete, M.; and Araque, A. 2009. Tripartite synapses: astrocytes process and control synaptic information. *Trends in neurosciences*, 32(8): 421–431.

Real, E.; Aggarwal, A.; Huang, Y.; and Le, Q. V. 2019. Regularized evolution for image classifier architecture search. In *Proceedings of the aaai conference on artificial intelligence*, volume 33, 4780–4789.

Ren, S.; He, K.; Girshick, R.; and Sun, J. 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28.

Rosenblatt, F. 1958. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6): 386.

Tian, Z.; Shen, C.; and Chen, H. 2020. Conditional convolutions for instance segmentation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16*, 282–298. Springer.

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

Wang, H.; Kembhavi, A.; Farhadi, A.; Yuille, A. L.; and Rastegari, M. 2019. Elastic: Improving cnns with dynamic scaling policies. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2258–2267.

Wang, R.; Cheng, M.; Chen, X.; Tang, X.; and Hsieh, C. 2021. Rethinking Architecture Selection in Differentiable NAS. In *Int. Conf. Learn. Represent.*

Wen, W.; Wu, C.; Wang, Y.; Chen, Y.; and Li, H. 2016. Learning structured sparsity in deep neural networks. *Advances in neural information processing systems*, 29: 2074–2082.

Xiao, H.; Wang, Z.; Zhu, Z.; Zhou, J.; and Lu, J. 2022. Shapley-NAS: Discovering Operation Contribution for Neural Architecture Search. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 11892–11901.

Xu, Y.; Wang, Y.; Han, K.; Tang, Y.; Jui, S.; Xu, C.; and Xu, C. 2021. Renas: Relativistic evaluation of neural architecture search. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4411–4420.

Yang, J.; Liu, Y.; and Xu, H. 2023. HOTNAS: Hierarchical Optimal Transport for Neural Architecture Search. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11990–12000.

Yang, Y.; Pan, L.; and Liu, L. 2023. Event Camera Data Pre-training. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 10699–10709.

Yu, H.; Li, H.; Hua, G.; Huang, G.; and Shi, H. 2023. Boosted Dynamic Neural Networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 10989–10997.

Zagoruyko, S.; and Komodakis, N. 2016. Wide Residual Networks. In *British Machine Vision Conference 2016*. British Machine Vision Association.

Zhang, L.; Chen, X.; Zhang, J.; Dong, R.; and Ma, K. 2022. Contrastive Deep Supervision. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXVI*, 1–19.

Zhao, B.; Cui, Q.; Song, R.; Qiu, Y.; and Liang, J. 2022. Decoupled knowledge distillation. In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*, 11953–11962.

Zhong, Z.; Yang, Z.; Deng, B.; Yan, J.; Wu, W.; Shao, J.; and Liu, C.-L. 2020. Blockqnn: Efficient block-wise neural network architecture generation. *IEEE transactions on pattern analysis and machine intelligence*, 43(7): 2314–2328.

Zoph, B.; Vasudevan, V.; Shlens, J.; and Le, Q. V. 2018. Learning transferable architectures for scalable image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 8697–8710.