Vision GNN (ViG)

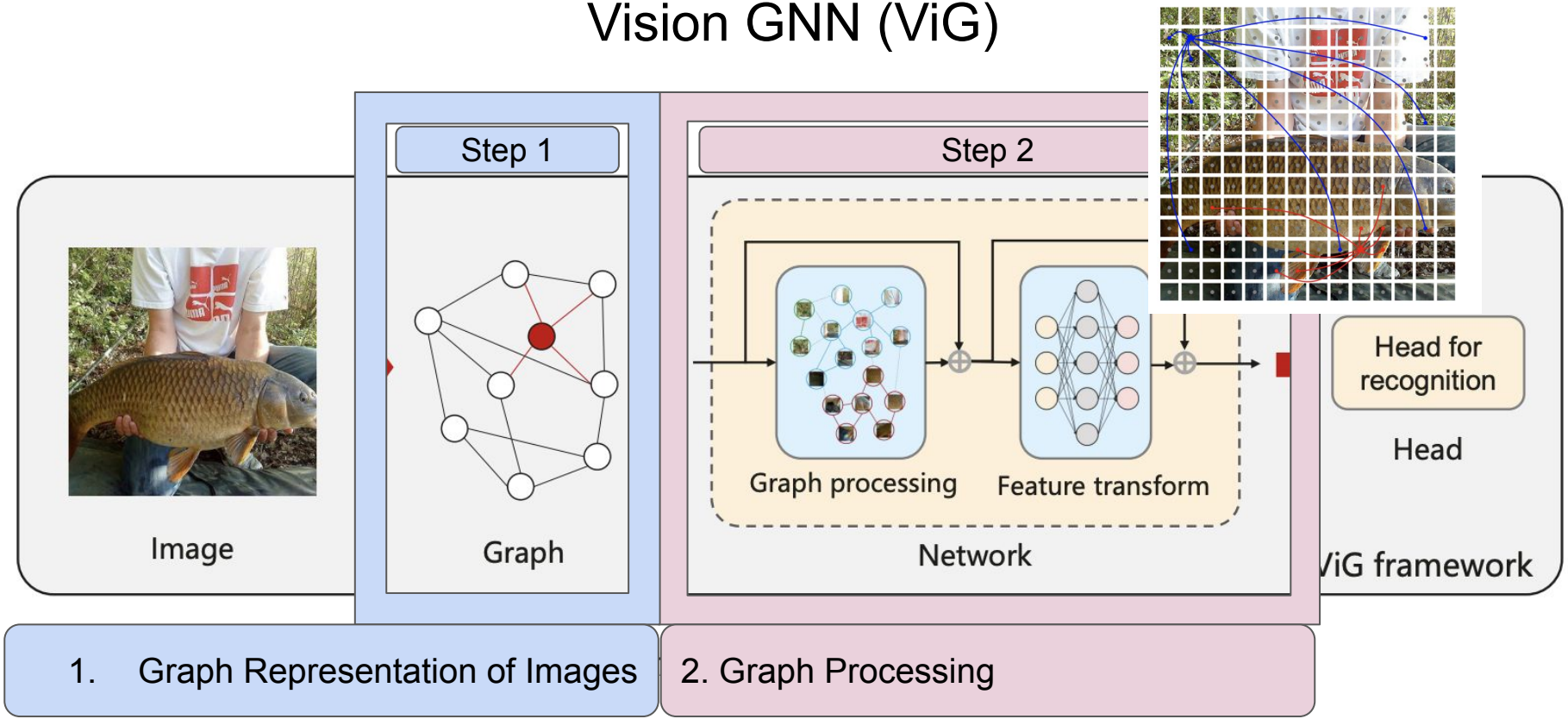
An Image is Worth Graph of Nodes

https://github.com/prodramp/DeepWorks/tree/main/ViG(VisionGNN)

https://arxiv.org/pdf/2206.00272v1.pdf
https://github.com/huawei-noah/CV-Backbones/

@avkashchauhan

Vision GNN (ViG)



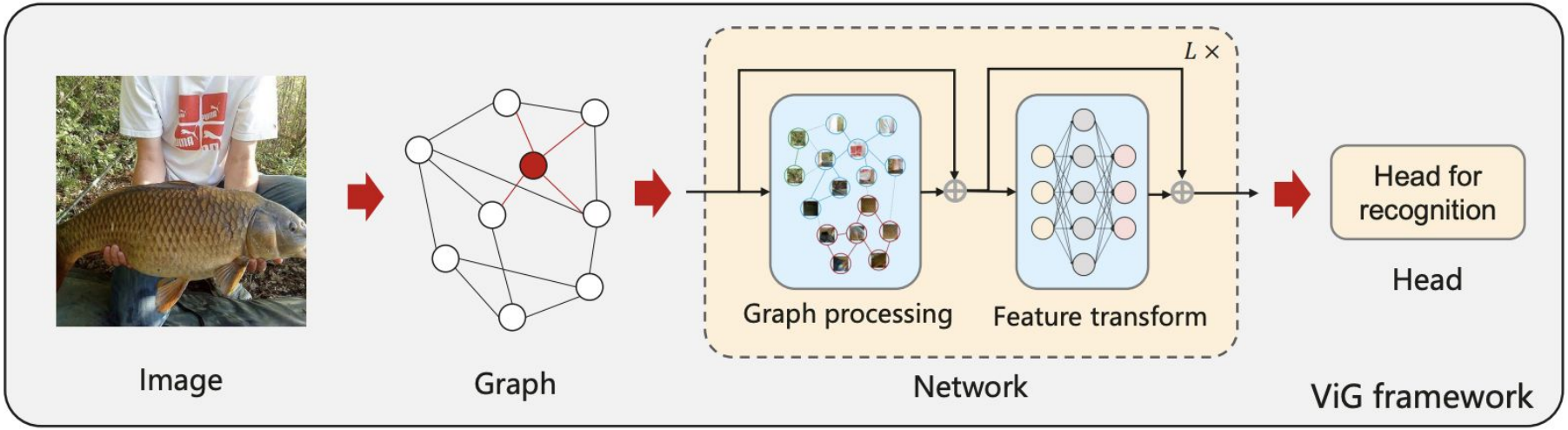


Figure 2: The framework of the proposed ViG model.

Source Image
[H x W x 3]

Image Patches
of [M x M]

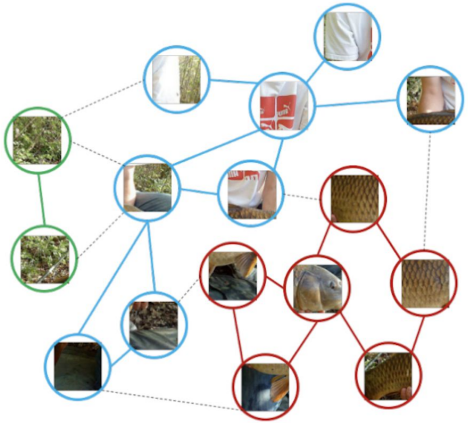
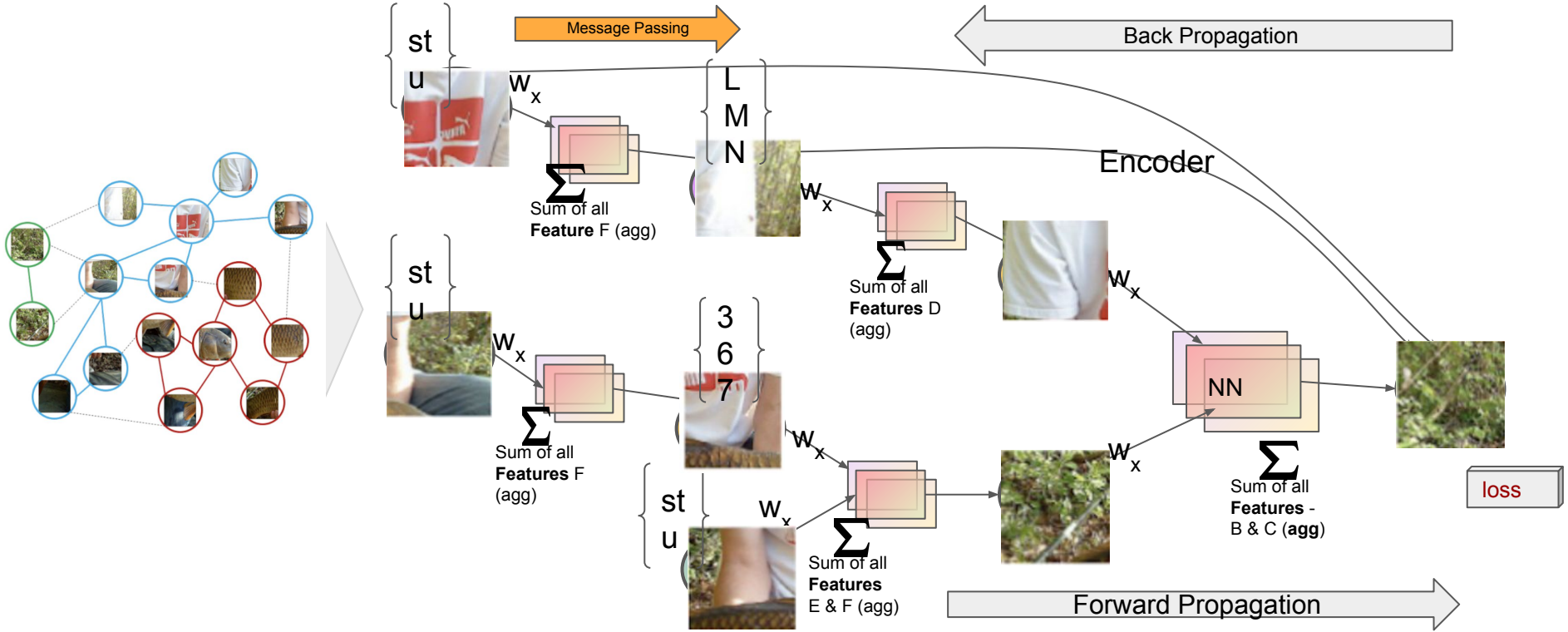


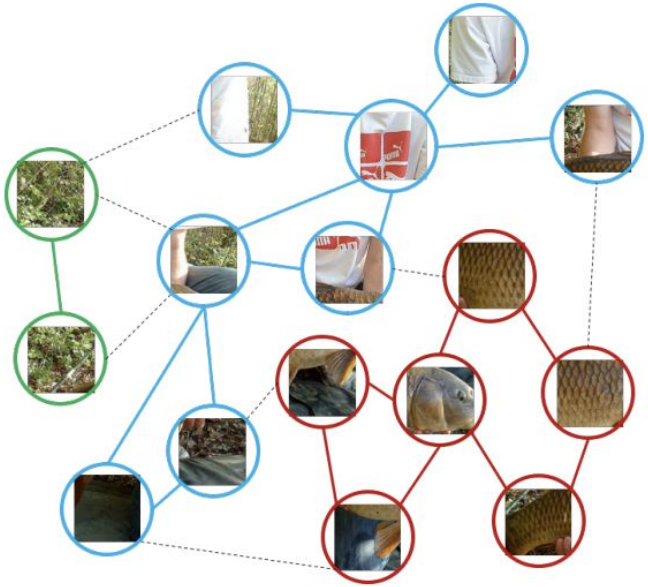
Image Graph



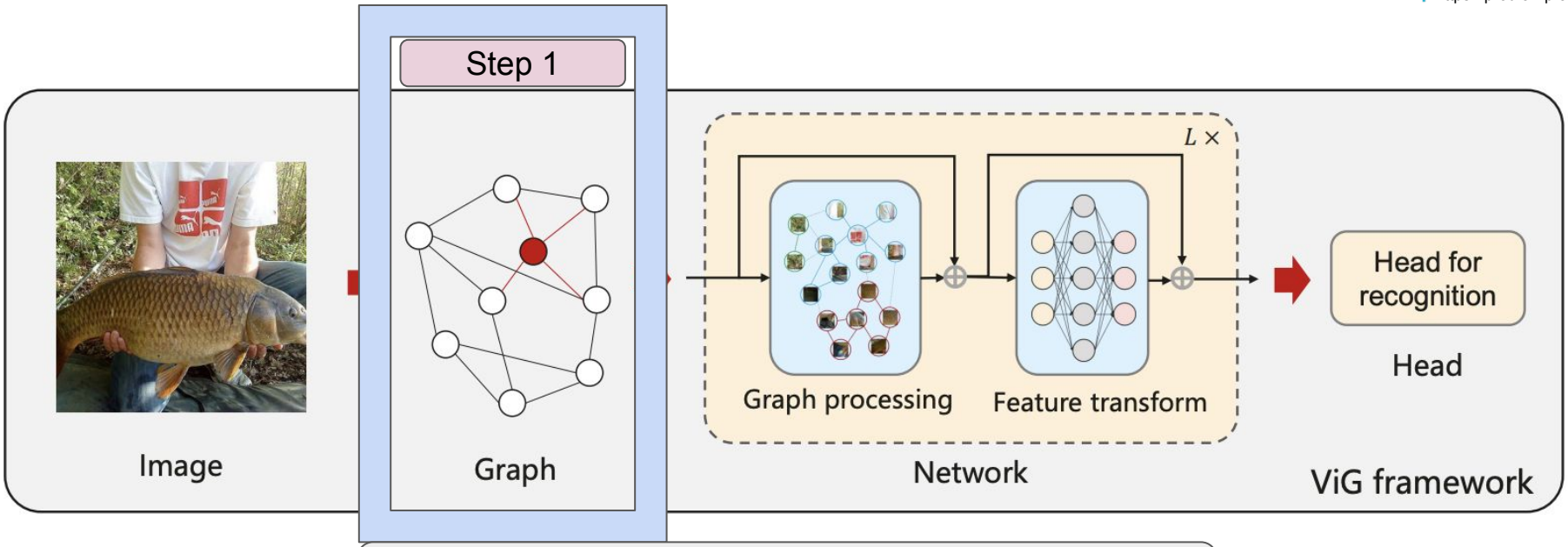
Repetition are done several times to improve Graph Neural Network

(a) Grid structure.

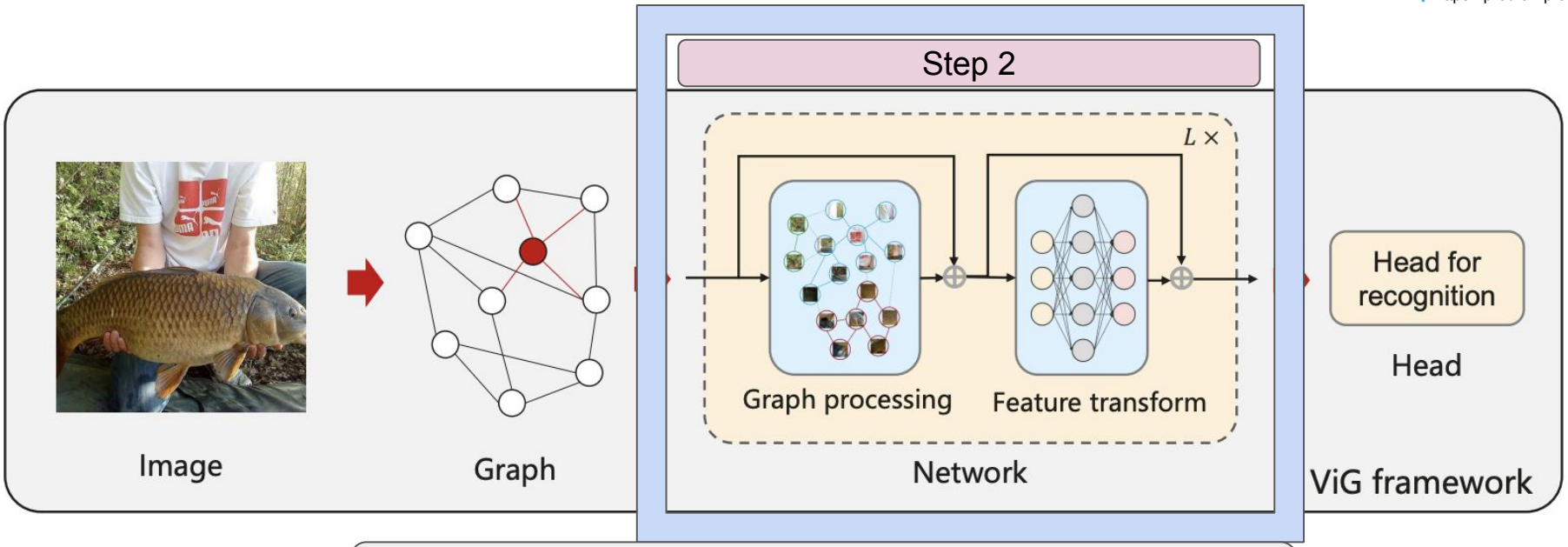
(b) Sequence structure.



(b) Graph structure.



1. Graph Representation of Images



2. Graph Processing

1. Graph Representation of Images

H x W x 3 = N Patches

Transforming Each patch into the feature Vector with D as Feature Dimension.

$Z = [\, x_1, x_2, x_3, \ldots\ldots x_n \,]$

Features are set of unordered nodes

$V = [\, v_1, v_2, v_3, \ldots\ldots v_n \,]$

1. For each node v_i its K nearest neighbors are found $N(v_i)$
2. Adding an edge e_{ji} directed from V_j to V_i for all $Vj \in N(v_i)$

Finally the Graph G is constructed:

$G = (V, E)$

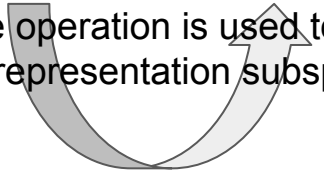
V all image patches
E - All edges

2. Graph Processing

GCN
Graph Convolution Network

A graph convolutional layer can exchange information between nodes by aggregating features from its neighbor nodes. Specifically, graph convolution operates as follows: G 0 = F(G, W)

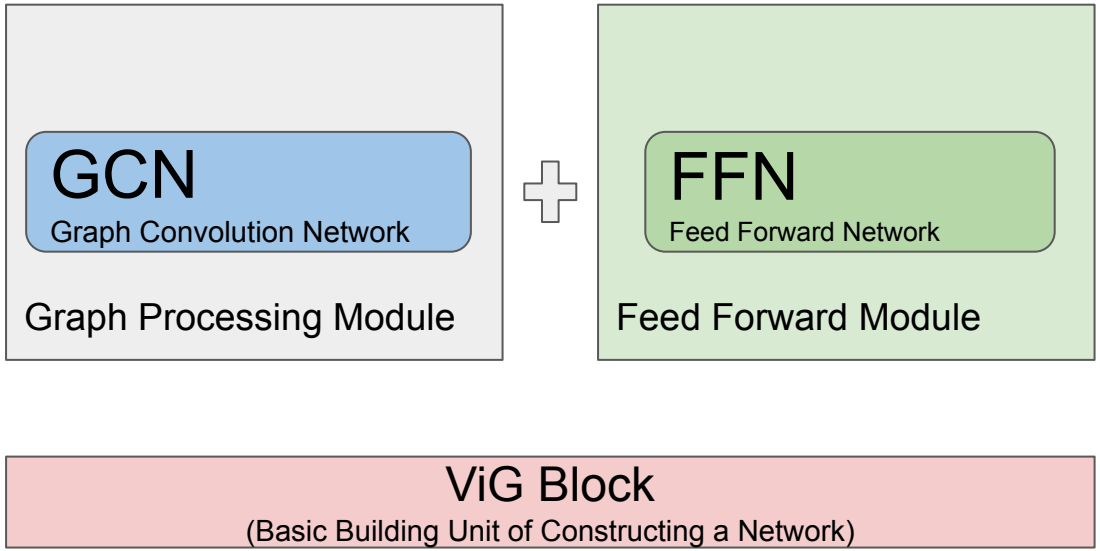
In graph convolution Multi-head update operation is used to allow the node to update information in multiple representation subspaces to benefit feature diversity.



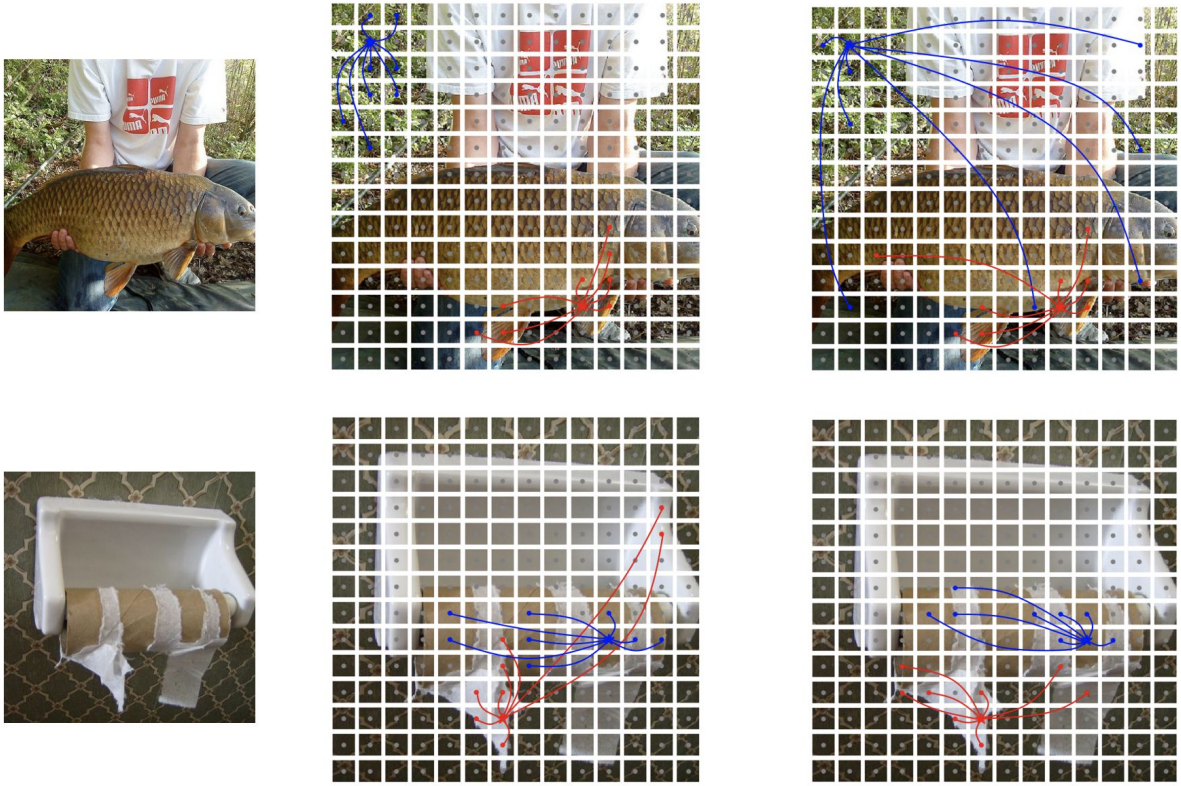
In GCN, repeatedly use of several graph convolution layers produces the graph data however over-smoothing feature of deep GCN, decreases the distinctiveness of node features and cause performance degradation for visual recognition.

FFN
Feed Forward Network

To encourage the feature transformation capacity and relief the over-smoothing phenomenon, a feed-forward network (FFN) is applied on each node. The FFN module is a simple multi-layer perceptron with two fully-connected layers:

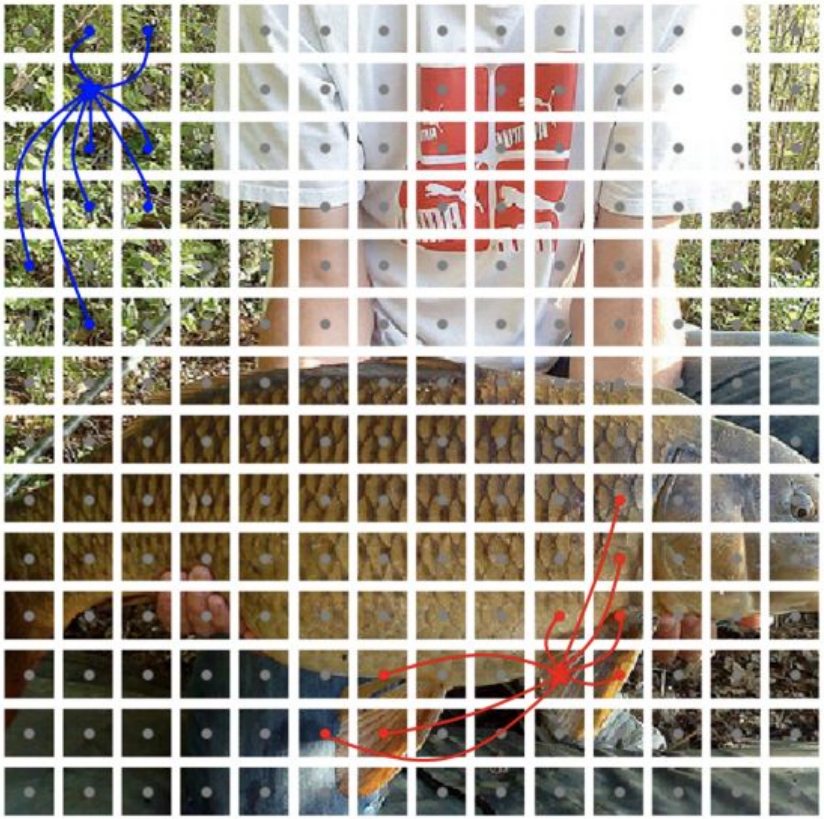


Constructed Graph Structure



(a) Input image. (b) Graph connection in the 1st block. (b) Graph connection in the 12th block.

Constructed Graph Structure



Graph connections in the 1st Block

Graph connections in the 12th Block

Computer Vision

(Commonly used Transformer)

Isotropic Architecture (ViT [8])

Input images features same size in whole network, size does not change

Pyramid Architecture ResNet [16]

Multiscale images properties and small spatial size as layer goes deep

Model	Resolution	Params (M)	FLOPs (B)	Top-1	Top-5
♠ ResMLP-S12 conv3x3 [48]	224×224	16.7	3.2	77.0	-
♠ ConvMixer-768/32 [50]	224×224	21.1	20.9	80.2	-
♠ ConvMixer-1536/20 [50]	224×224	51.6	51.4	81.4	-
♦ ViT-B/16 [8]	384×384	86.4	55.5	77.9	-
♦ DeiT-Ti [49]	224×224	5.7	1.3	72.2	91.1
♦ DeiT-S [49]	224×224	22.1	4.6	79.8	95.0
♦ DeiT-B [49]	224×224	86.4	17.6	81.8	95.7
■ ResMLP-S24 [48]	224×224	30	6.0	79.4	94.5
■ ResMLP-B24 [48]	224×224	116	23.0	81.0	95.0
■ Mixer-B/16 [47]	224×224	59	11.7	76.4	-
★ ViG-Ti (ours)	224×224	7.1	1.3	**73.9**	**92.0**
★ ViG-S (ours)	224×224	22.7	4.5	**80.4**	**95.2**
★ ViG-B (ours)	224×224	86.8	17.7	**82.3**	**95.9**

Model	Resolution	Params (M)	FLOPs (B)	Top-1	Top-5
♠ ResNet-18 [16, 56]	224×224	12	1.8	70.6	89.7
♠ ResNet-50 [16, 56]	224×224	25.6	4.1	79.8	95.0
♠ ResNet-152 [16, 56]	224×224	60.2	11.5	81.8	95.9
♠ BoTNet-T3 [44]	224×224	33.5	7.3	81.7	-
♠ BoTNet-T3 [44]	224×224	54.7	10.9	82.8	-
♠ BoTNet-T3 [44]	256×256	75.1	19.3	83.5	-
♦ PVT-Tiny [54]	224×224	13.2	1.9	75.1	-
♦ PVT-Small [54]	224×224	24.5	3.8	79.8	-
♦ PVT-Medium [54]	224×224	44.2	6.7	81.2	-
♦ PVT-Large [54]	224×224	61.4	9.8	81.7	-
♦ CvT-13 [57]	224×224	20	4.5	81.6	-
♦ CvT-21 [57]	224×224	32	7.1	82.5	-
♦ CvT-21 [57]	384×384	32	24.9	83.3	-
♦ Swin-T [33]	224×224	29	4.5	81.3	95.5
♦ Swin-S [33]	224×224	50	8.7	83.0	96.2
♦ Swin-B [33]	224×224	88	15.4	83.5	96.5
■ CycleMLP-B2 [4]	224×224	27	3.9	81.6	-
■ CycleMLP-B3 [4]	224×224	38	6.9	82.4	-
■ CycleMLP-B4 [4]	224×224	52	10.1	83.0	-
■ Poolformer-S12 [64]	224×224	12	2.0	77.2	93.5
■ Poolformer-S36 [64]	224×224	31	5.2	81.4	95.5
■ Poolformer-M48 [64]	224×224	73	11.9	82.5	96.0
★ Pyramid ViG-Ti (ours)	224×224	10.7	1.7	**78.2**	**94.2**
★ Pyramid ViG-S (ours)	224×224	27.3	4.6	**82.1**	**96.0**
★ Pyramid ViG-M (ours)	224×224	51.7	8.9	**83.1**	**96.4**
★ Pyramid ViG-B (ours)	224×224	92.6	16.8	**83.7**	**96.5**