ViG-SP: A Segmentation-Driven Approach to Image Classification

Supervised by:
Dr. Karolis Martinkus
Ard Kastrati
Prof. Dr. Roger Wattenhofer

Philip Toma
Mateo Diaz-Bone
Stefan Scholbe

Revisiting ResNets: Improved Training and Scaling Strategies

Irwan Bello Google Brain
William Fedus Google Brain
Xianzhi Du Google Brain
Ekin D. Cubuk Google Brain
Aravind Srinivas UC Berkeley
Tsung-Yi Lin Google Brain
Jonathon Shlens Google Brain
Barret Zoph Google Brain

Approaches to Image Classification

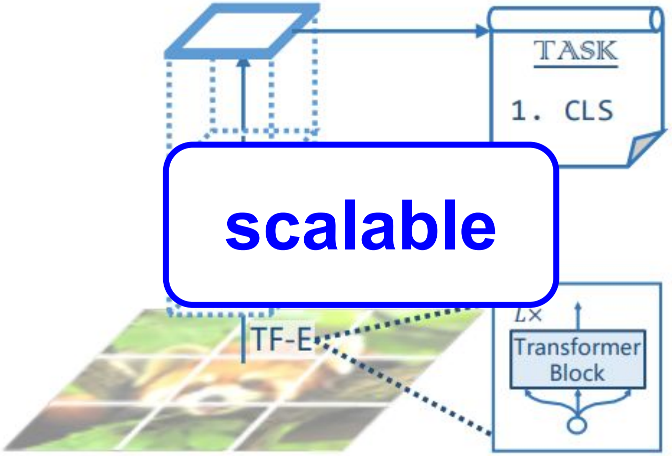
AN IMAGE IS WORTH 16X16 WORDS: TRANSFORMERS FOR IMAGE RECOGNITION AT SCALE

Alexey Dosovitskiy[*,†], **Lucas Beyer**[*], **Alexander Kolesnikov**[*], **Dirk Weissenborn**[*], **Xiaohua Zhai**[*], **Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, Neil Houlsby**[*,†]
[*]equal technical contribution, [†]equal advising
Google Research, Brain Team
{adosovitskiy, neilhoulsby}@google.com

Approaches to Image Classification

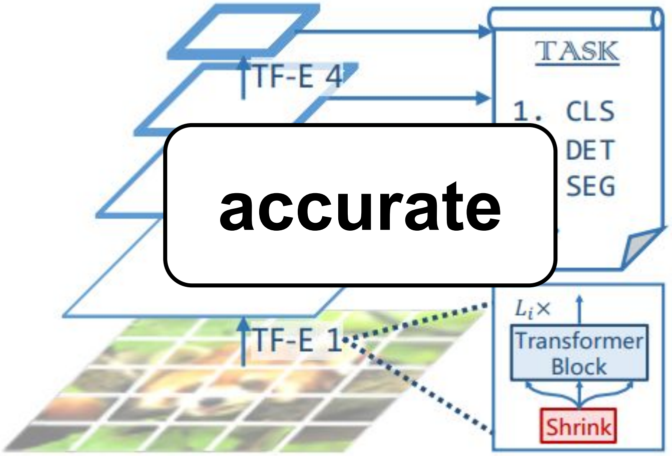
Isotropic

- Consistent feature size
- Uniform receptive fields
- Efficient and compact design

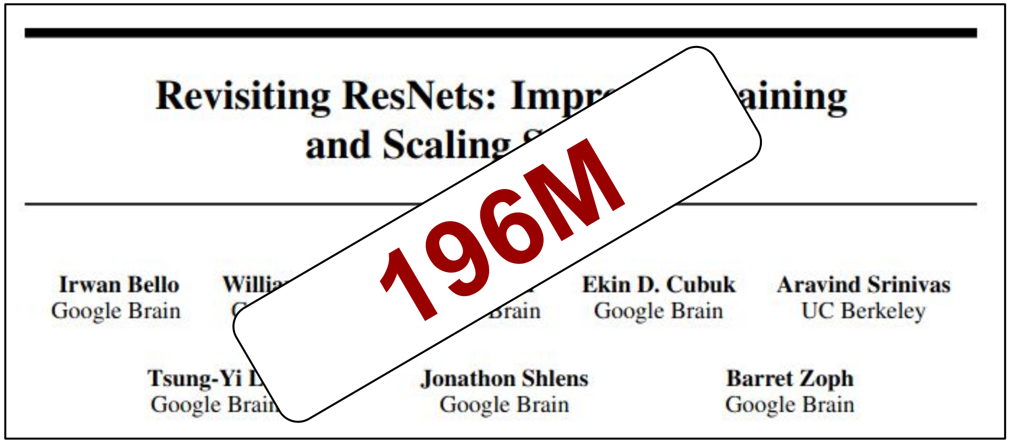


Pyramid

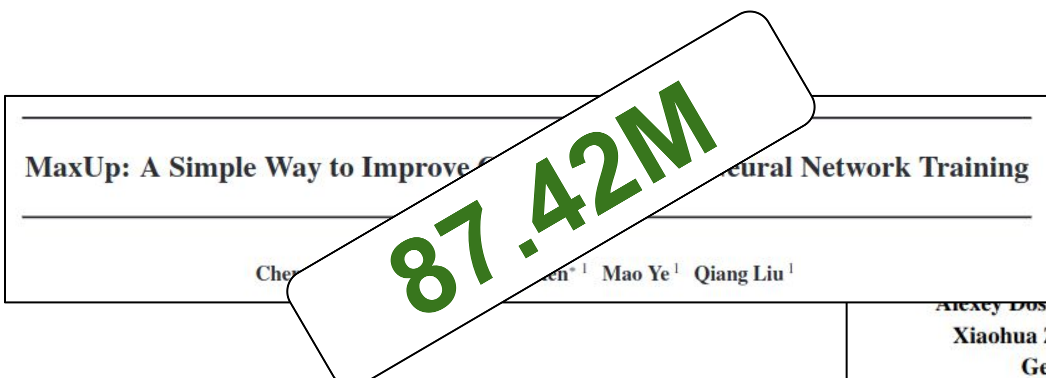
- Multi-layered, hierarchical
- Diverse receptive fields
- Captures fine to broad features



Wang et al. 2021



Approaches to Image Classification



Graph-Based Approach to Image Classification

Vision GNN: An Image is Worth Graph of Nodes

Kai Han[1,2*] Yunhe Wang[2*] Jianyuan Guo[2] Yehui Tang[2,3] Enhua Wu[1,4]
[1]State Key Lab of Computer Science, ISCAS & UCAS
[2]Huawei Noah's Ark Lab
[3]Peking University [4]University of Macau
{kai.han,yunhe.wang}@huawei.com, weh@ios.ac.cn

Published @ NeurIPS 2022 [3]

Vision GNN

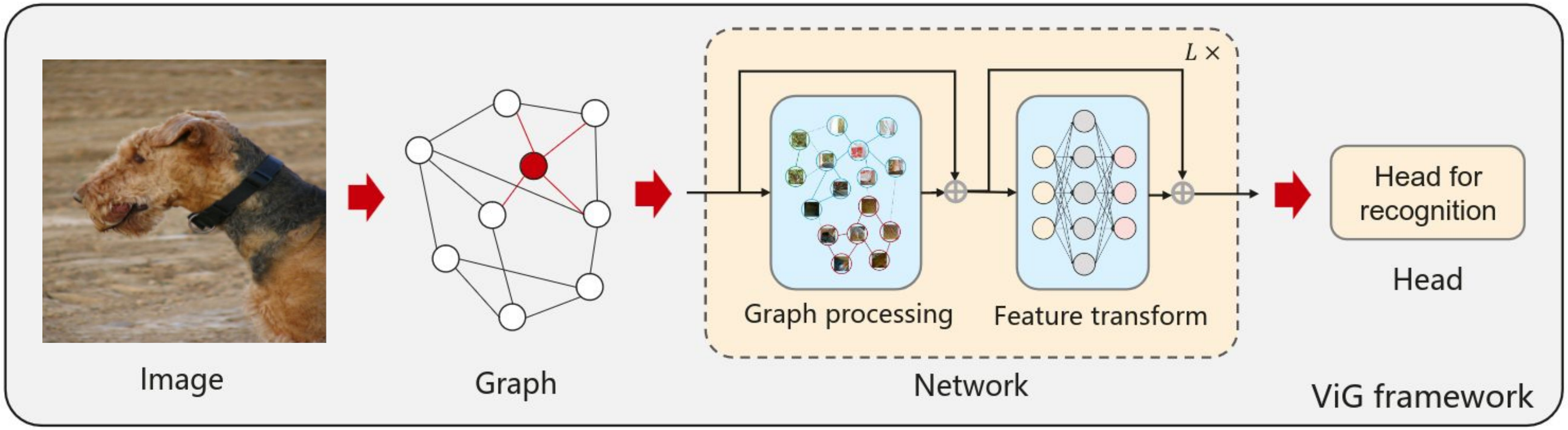
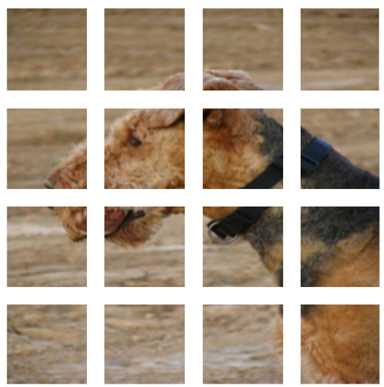


Image
Graph
$L\times$
Graph processing
Feature transform
Network
Head for recognition
Head
ViG framework

Graph Representation Inference

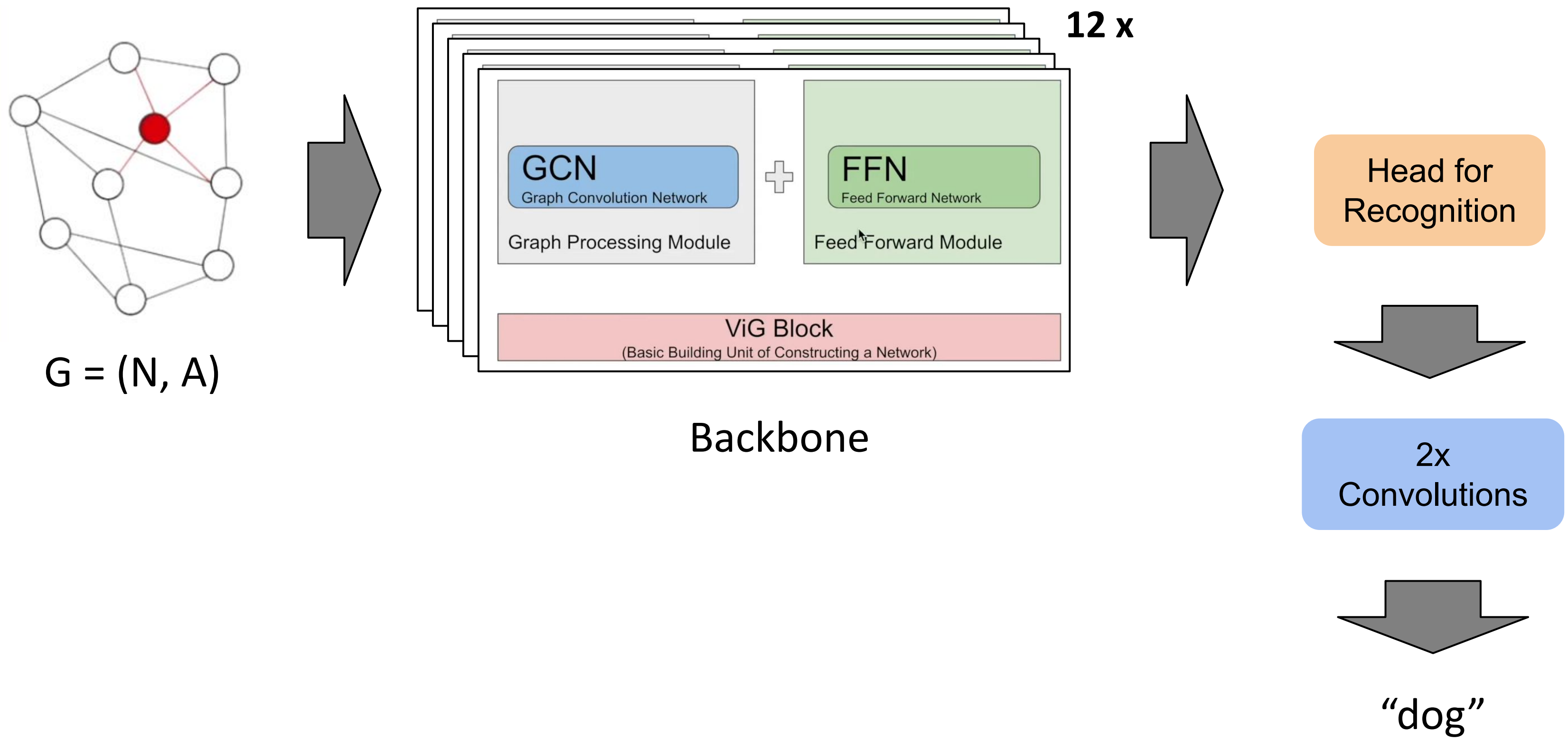
14 x 14



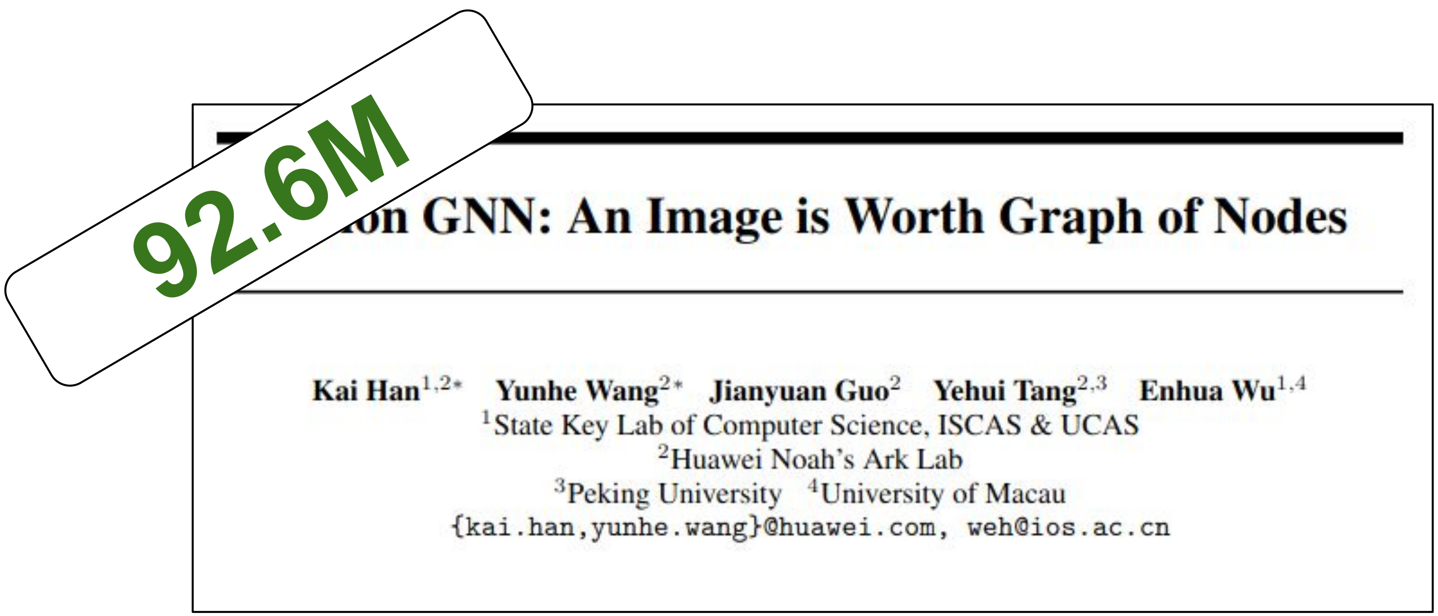
K-Nearest Neighbors for each patch

G = (N, A)

Graph Updates, Feature Transformations & Prediction



Graph-Based Approach to Image Classification


92.6M

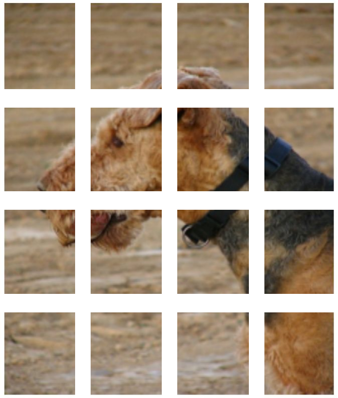
on GNN: An Image is Worth Graph of Nodes

Kai Han[1,2*] Yunhe Wang[2*] Jianyuan Guo[2] Yehui Tang[2,3] Enhua Wu[1,4]
[1]State Key Lab of Computer Science, ISCAS & UCAS
[2]Huawei Noah's Ark Lab
[3]Peking University [4]University of Macau
{kai.han,yunhe.wang}@huawei.com, weh@ios.ac.cn

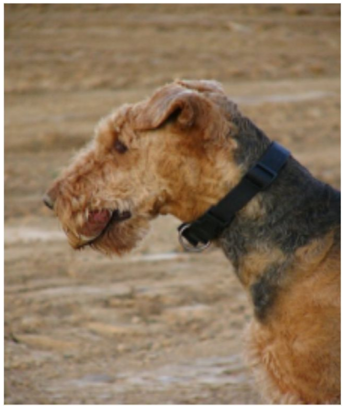
Published @ NeurIPS 2022

Combining ViG with Image Segmentation

Image Segmentations



Grid segmentation

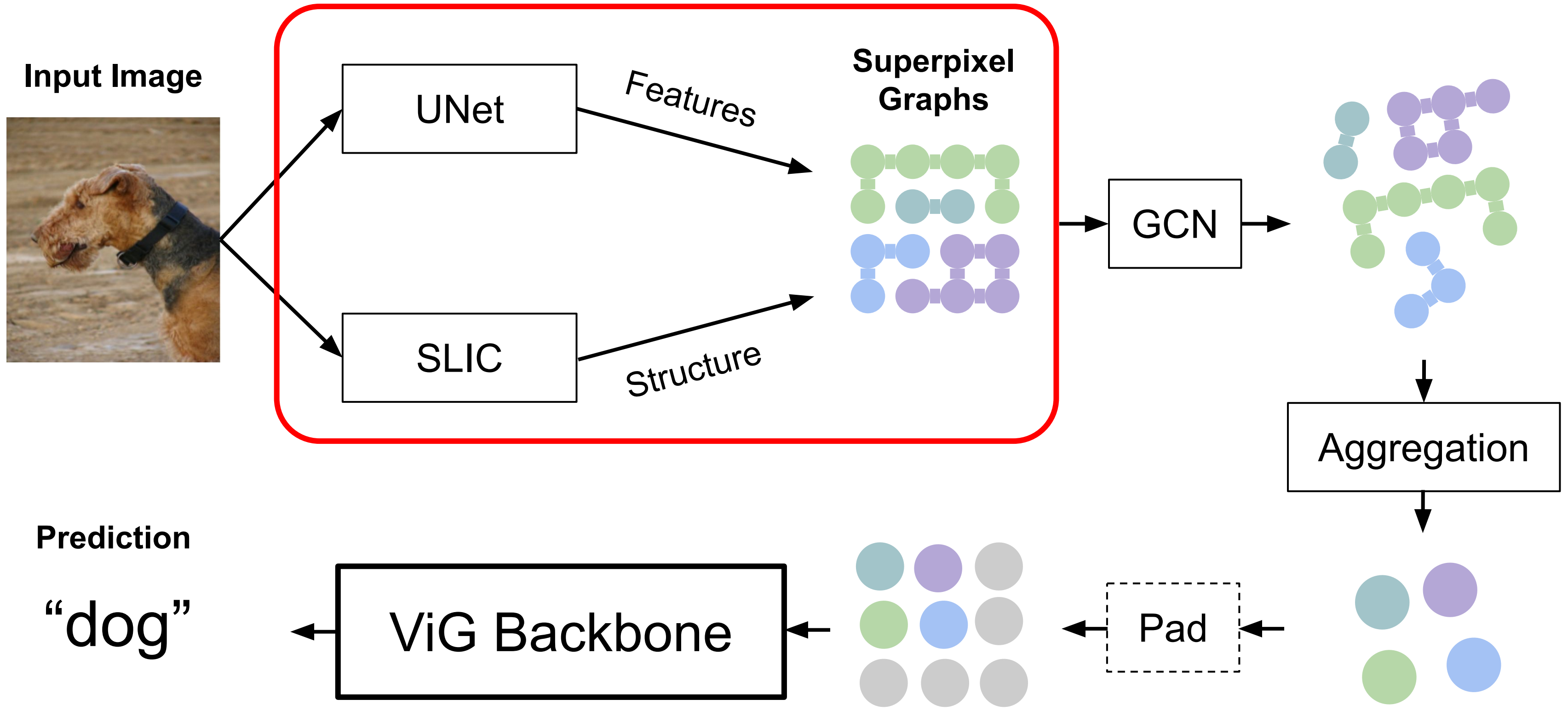


Original Image

SLIC segmentation

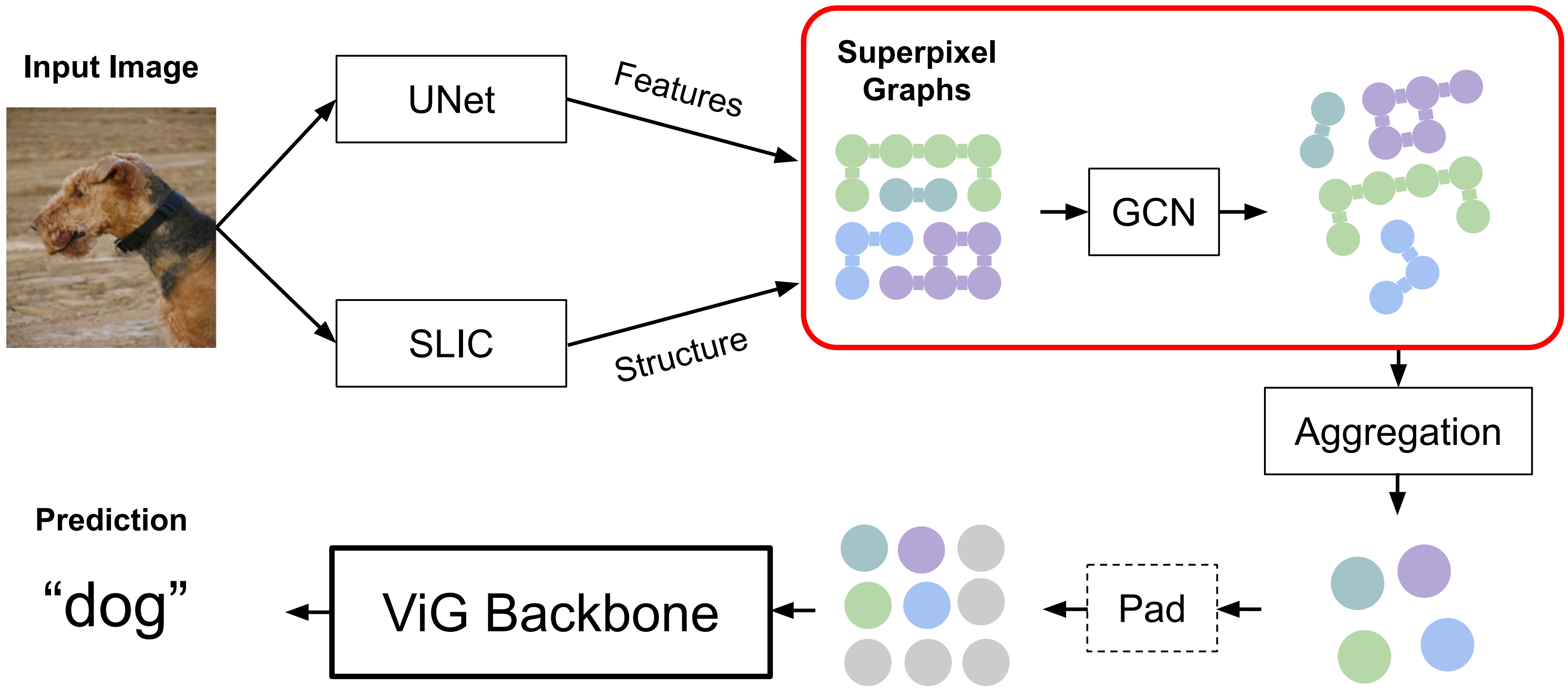
- Dividing Images into rectangular patches might not be optimal
- SLIC Segmentation covers irregular structures encaptured in the image

Incorporating Image Segmentation into VisionGNN



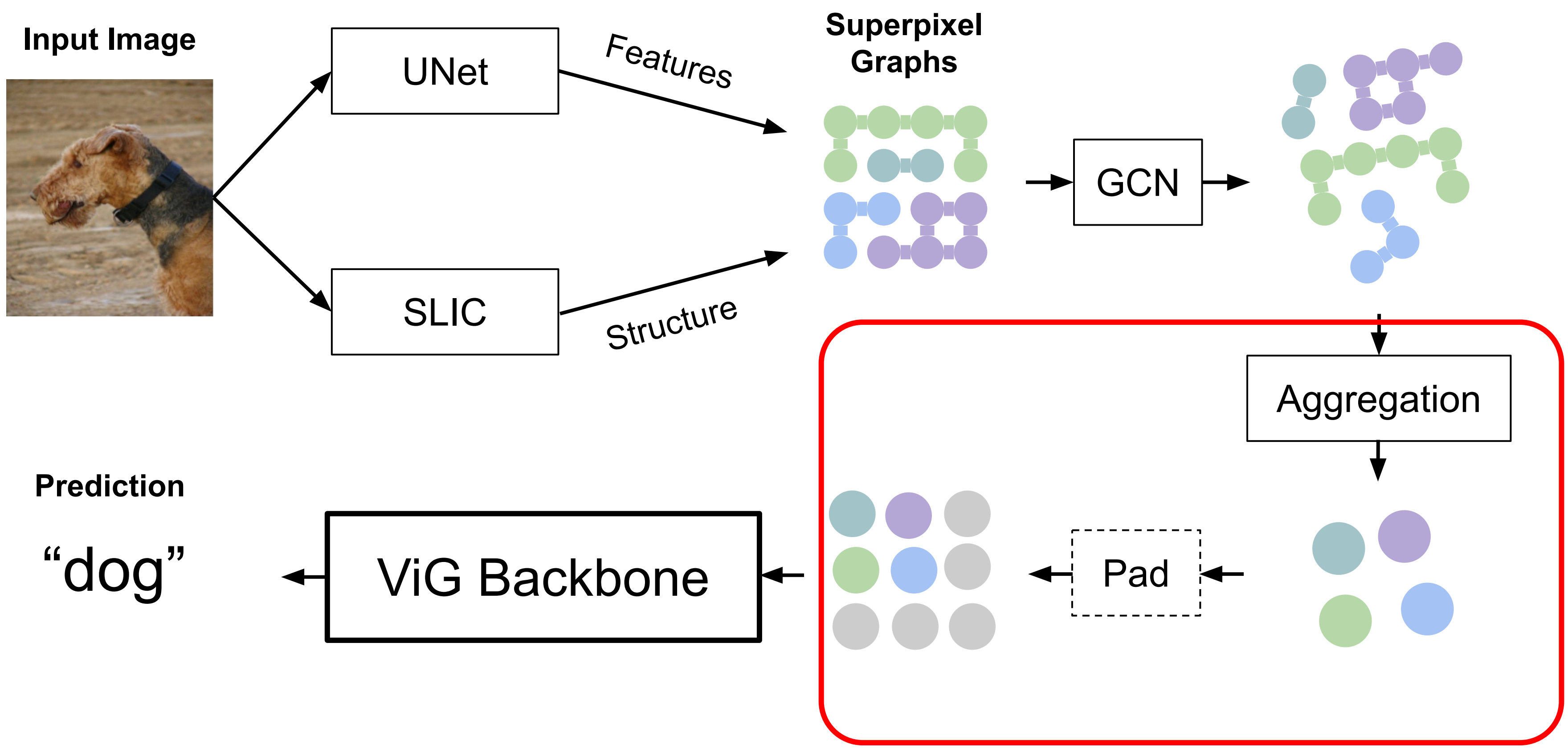
- SLIC segmentation provides graph structure
- UNet model provides per-pixel features

Incorporating Image Segmentation into VisionGNN



- On each superpixel induced graph a custom GCN enables sharing of node information

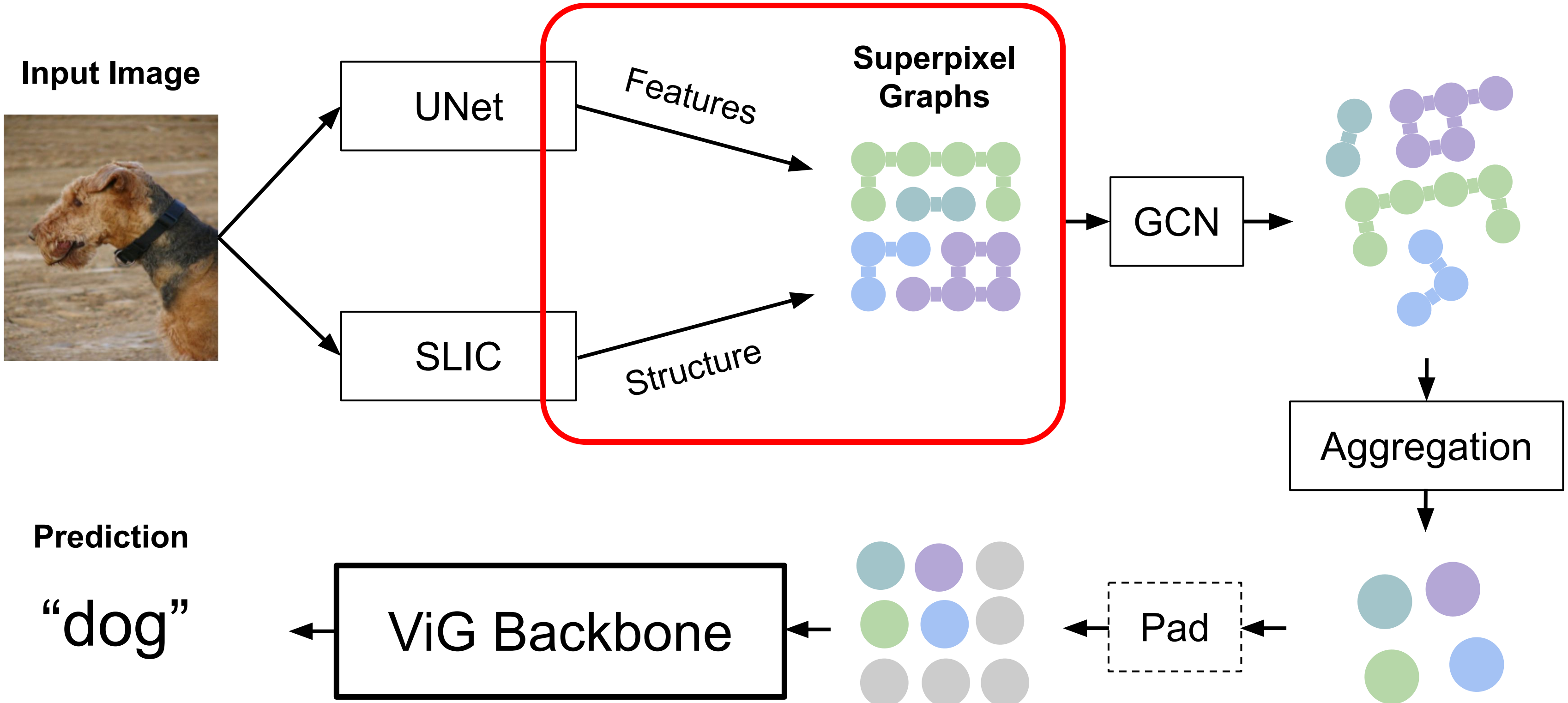
Incorporating Image Segmentation into VisionGNN



- After GCN we perform a simple aggregation to arrive at one feature vector per superpixel
- After padding we forward the features to the original ViG Backbone

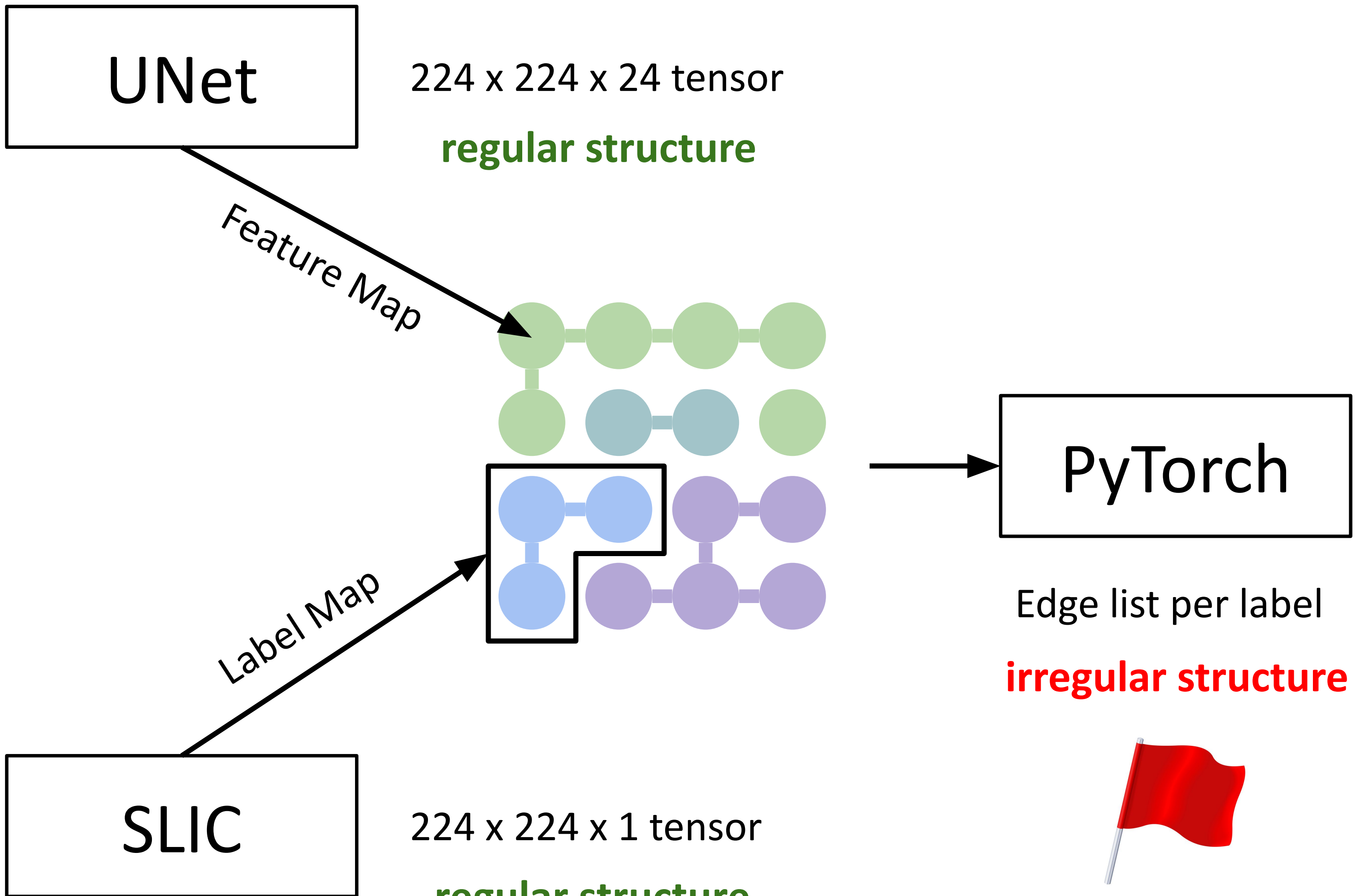
Optimizations

Optimizations



- Unique to our architecture: No optimized PyTorch function available.
- On the "hot path" – every millisecond counts!

Optimizations: Generating Superpixel Graphs



Optimizations: Attempts

Attempt 1:

Precompute graphs and optimize for fast disk access.

➢ Augmentations can’t be precomputed

Attempt 2:

Generate graphs on the fly.

➢ Needs to be extremely fast

Optimizations: Implementations

Python
Simple double for-loop
249 ms

Python + NumPy
Vectorization through index transformations
35 ms

C/C++ Binding
Low-level and memory optimizations
2 ms

~ 20% time save per epoch

Results

Semester Project

From Pixels to Nodes: A Segmentation-Driven Approach to Image Classification

Mateo Diaz-Bone, Philip Toma, Stefan Scholbe

Ablation Study

	ViG-Ti	ViG-B	ViG-SP196	ViG-SP100	ViG-SP-Grid
Top-5 (%)	96.33	97.10	***97.78***	97.28	97.44
Top-1 (%)	83.06	86.46	***87.37***	86.22	86.12

Ablation Study

	ViG-Ti	ViG-B	ViG-SP196	ViG-SP100	ViG-SP-Grid
Top-5 (%)	96.33	97.10	***97.78***	97.28	97.44
Top-1 (%)	83.06	86.46	***87.37***	86.22	86.12

Result:

On ImageNet-100, the ViG-SP196 model outperformed all other configurations

➤ SLIC with approx. 196 segments

Ablation Study

	ViG-Ti	ViG-B	ViG-SP196	ViG-SP100	ViG-SP-Grid
Top-5 (%)	96.33	97.10	***97.78***	97.28	97.44
Top-1 (%)	83.06	86.46	***87.37***	86.22	86.12

Result:

On ImageNet-100, the ViG-SP196 model outperformed all other configurations

➤ SLIC with approx. 196 segments

➤ Chosen for the benchmark on ImageNet

Ablation Study

	ViG-Ti	ViG-B	ViG-SP196	ViG-SP100	ViG-SP-Grid
Top-5 (%)	96.33	97.10	***97.78***	97.28	97.44
Top-1 (%)	83.06	86.46	***87.37***	86.22	86.12

Result:

On ImageNet-100, the ViG-SP196 model outperformed all other configurations

➤ SLIC with approx. 196 segments

➤ Chosen for the benchmark on ImageNet

Baseline

	ViG-Ti
Top-5 (%)	92.0
Top-1 (%)	73.9

- Least accurate ViG model
- Smallest ViG model
- Underlying backbone of our architecture

Benchmark

	Validation-Set	Test-Set
Top-5 (%)	88.38	87.67
Top-1 (%)	68.26	67.37

- ViG-parameters in the backbone
- 3 weeks training time
- 130 epochs
- Small batch size (h/w constraints)

Results

37M

Semester Project

Combining Graph- and Convolutional Neural Networks and Image Segmentation to Improve Image Classification

Mateo Diaz-Bone, Philip Toma, Stefan Scholbe

Conclusion

Positive:

Graph representations of images are more flexible

➤ Ability to learn on irregular structures and patterns

Good results in combination with CNNs

Negative:

High latency in training on graph representations and graph creation

➤ Difficult computational challenge

Thanks for your attention!

And the possibility to do this project.

Supervised by:
Dr. Karolis Martinkus
Ard Kastrati
Prof. Dr. Roger Wattenhofer

Philip Toma
Mateo Diaz-Bone
Stefan Scholbe