Computer Vision
2024/2025

Valerio Massimo Carioti

Efficient Anomaly Detection in Industrial Images using Transformers with Dynamic Tanh

The Challenge:

Developing an Efficient Anomaly Detection System

- Anomaly detection is required to guarantee high quality products
- Vision Transformers are computationally demanding
- Dynamic Tanh offers an easy way to increase efficiency

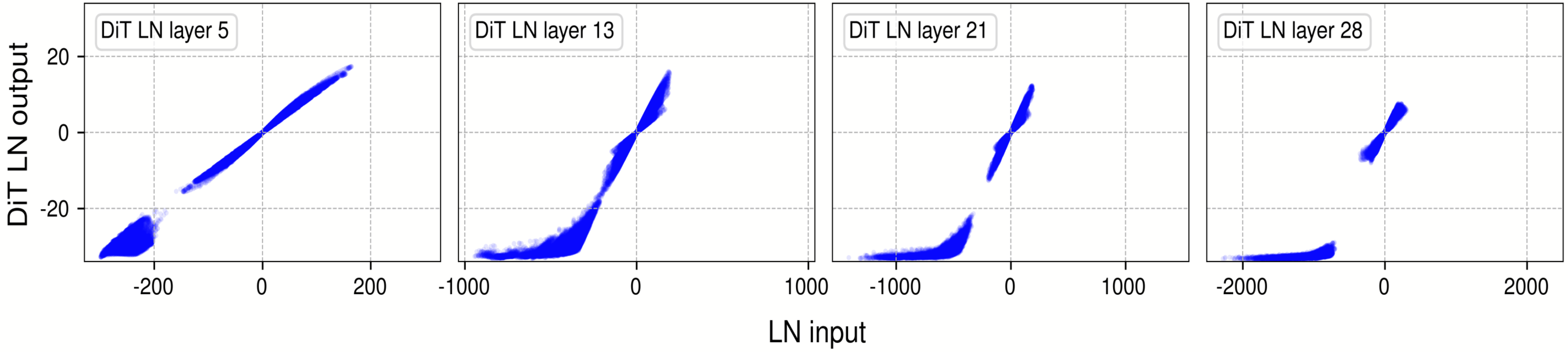
Tested Methods

1 – Better implementation of a model

- Use of VT-ADL[1]
- Full re-implementation to have more hyperparameters available

2 - Use of Dynamic Tanh[2]

- This layer offers a faster alternative to layer normalization in transformers



[1] Mishra P., Verk R., Fornasier D., Piciarelli C., Foresti G. L., *VT-ADL: A Vision Transformer Network for Image Anomaly Detection and Localization* (2021)

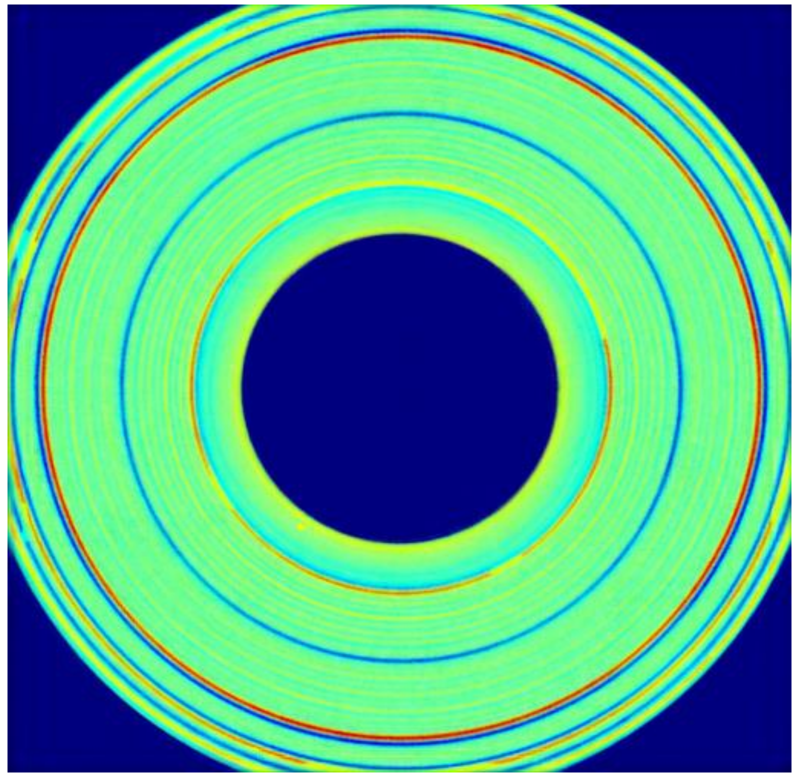
[2] Zhu J., Chen X., He K., LeCun Y., Liu Z., *Transformers without Normalization* (2025)

The Model

VT-ADL

Structure:
- ViT encoder
- Add of noise
- Routing algorithm
- Decoder recontruction
- Mixture Density Network
- Upscaling + smoothing

Loss:
Composed by:
- MDN loss
- SSIM loss
- Reconstruction loss (MAE)



VT-ADL re-implementation

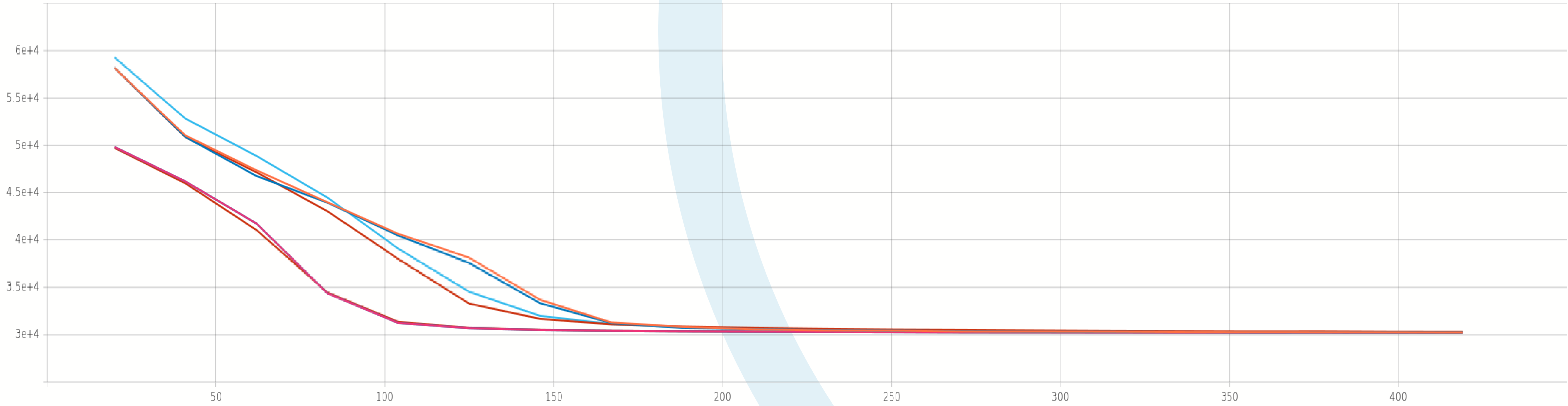
Available efficiency-related hyperparameters:

Original Implementation:
- Routing iterations
- Number of gaussians
- Encoder depth
- Encoder feed forward dimension

Our Implementation:
- Image size
- Patch size
- Latent channels
- Encoder heads
- Encoder depth
- Encoder feed forward dimension
- Capsules per patch
- Capsule size
- Routing iterations
- Number of gaussians
- Use of dynamic tanh
- Matrix multiplication precision

Exploring Hyperparameters

- Changed one hyperparameter at the time
- Trained for 20 epochs
- Kept the new value if the cumulative loss was less than or almost equal to the previous one



Loss evolution on MVTech with different hyperparameters

VT-ADL re-implementation

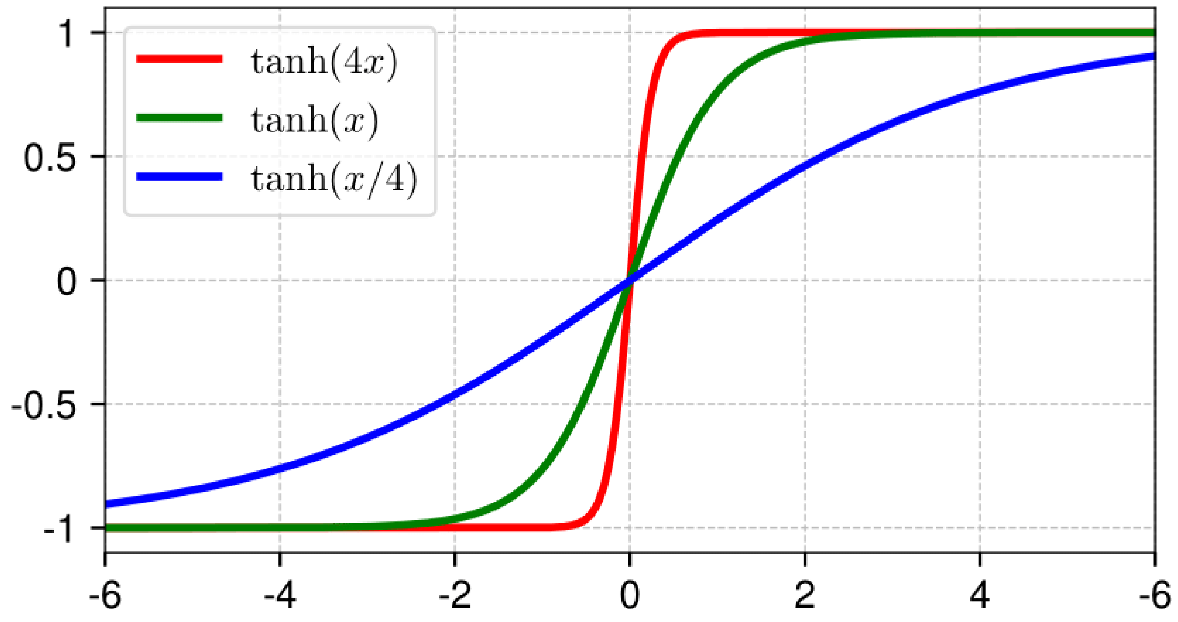
What achieved:

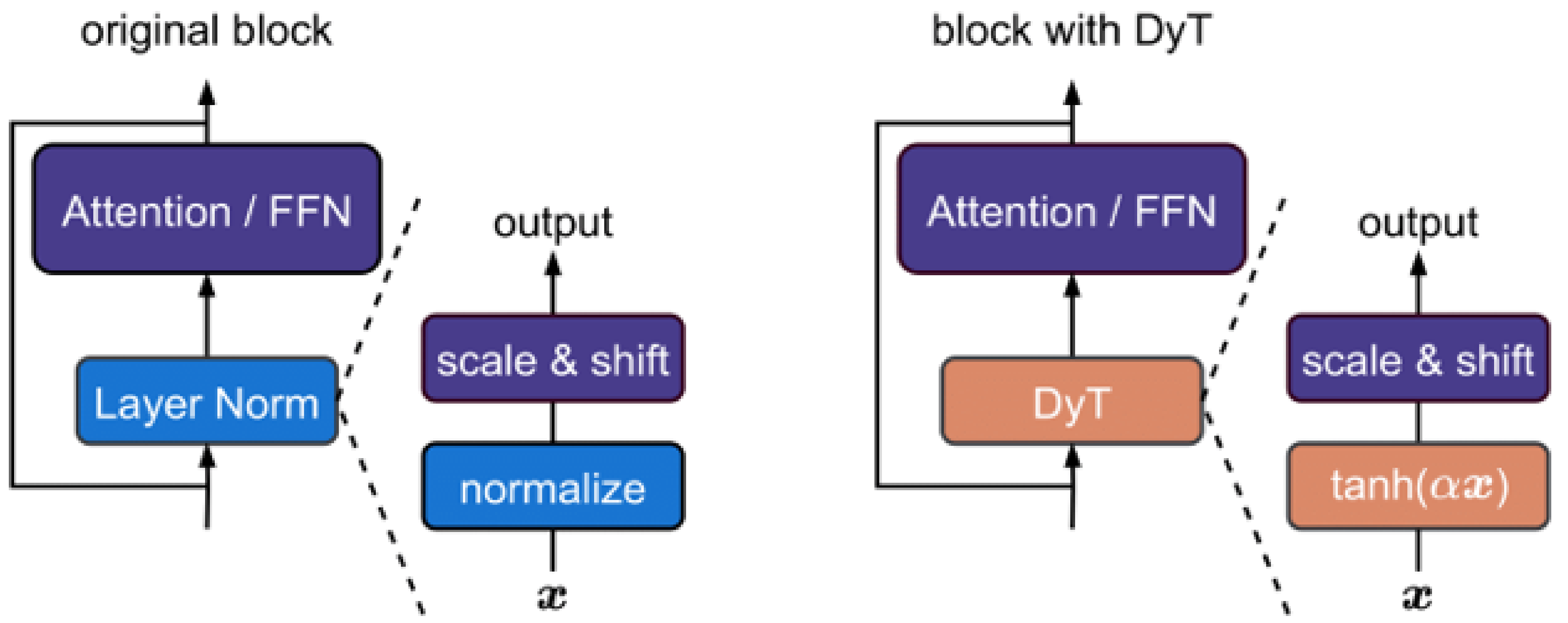
Changed Parameter	Original Value	New MVTech Value	New BTAD Value
Encoder Heads	8	8	2
Encoder depth	6	2	2
Encoder feed forward dimension	1024	512	1024
Capsules per patch	64	32	64
Capsule dimension	8	8	2
Routing iterations	3	1	1
Number of gaussians	150	50	50
Precision	Full	bf16-mixed	bf16-mixed

1
2
3
4
5
6
7
8
9
10
11
12
13

Use of Dynamic Tanh

- Replaces the LayerNorm in the transformer encoder
- Initialized with alpha = 0.5

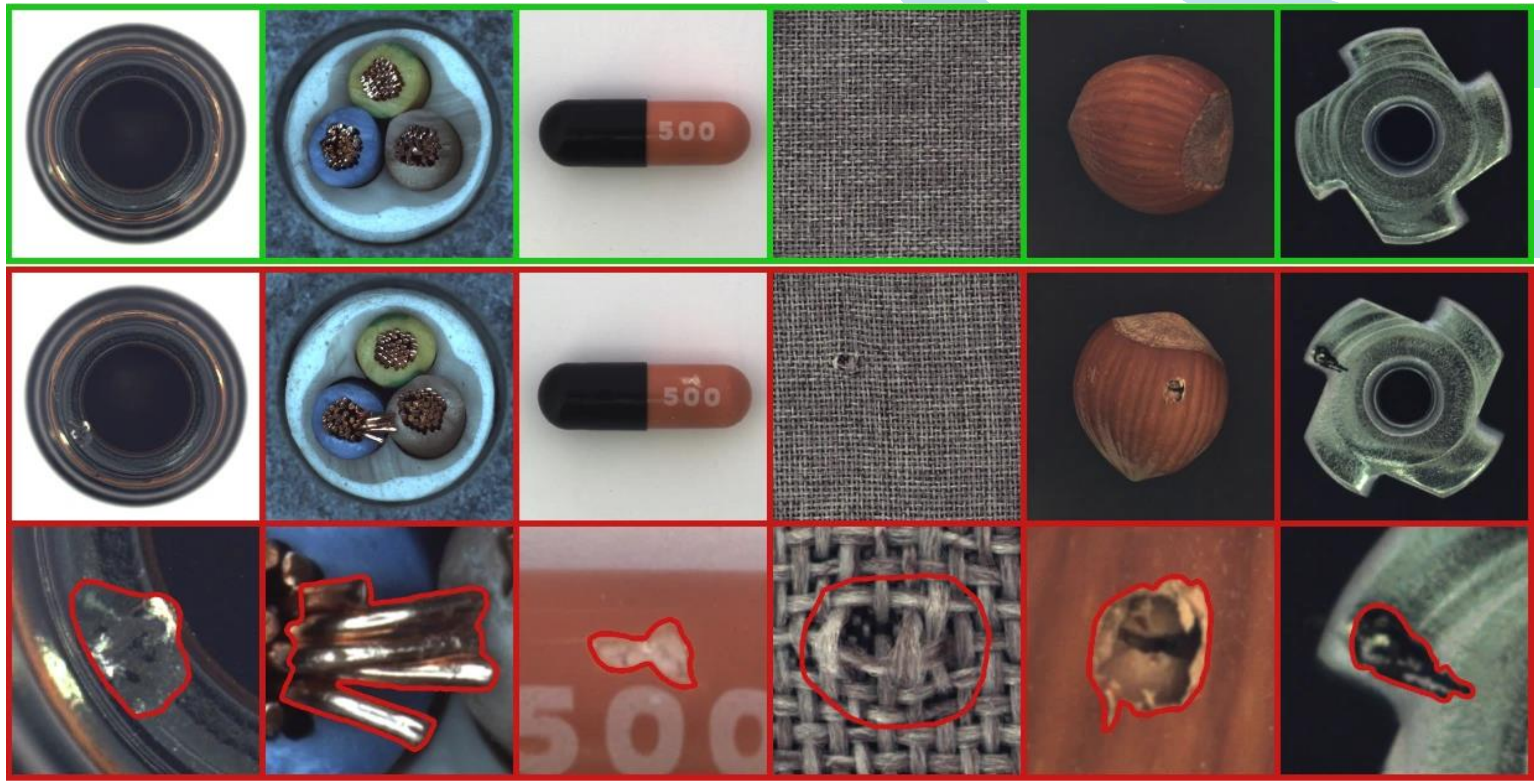




The Datasets

2 datasets have been used:

- MVTech[3]: a benchmark dataset focused on industrial inspection
- BTAD: an industrial anomaly dataset with real data from the beanTech company



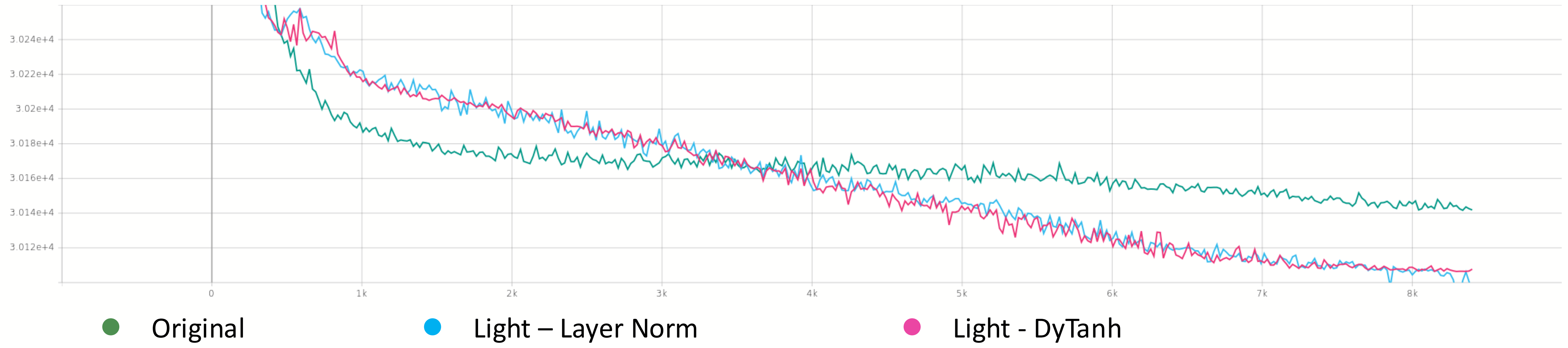
[3] Bergmann P., Fauser M., Sattlegger D., Steger C., *MVTec AD — A Comprehensive Real-World Dataset for Unsupervised Anomaly Detection* (2021)

1
2
3
4
5
6
7
8
9
10
11
12
13

Results

MVTech

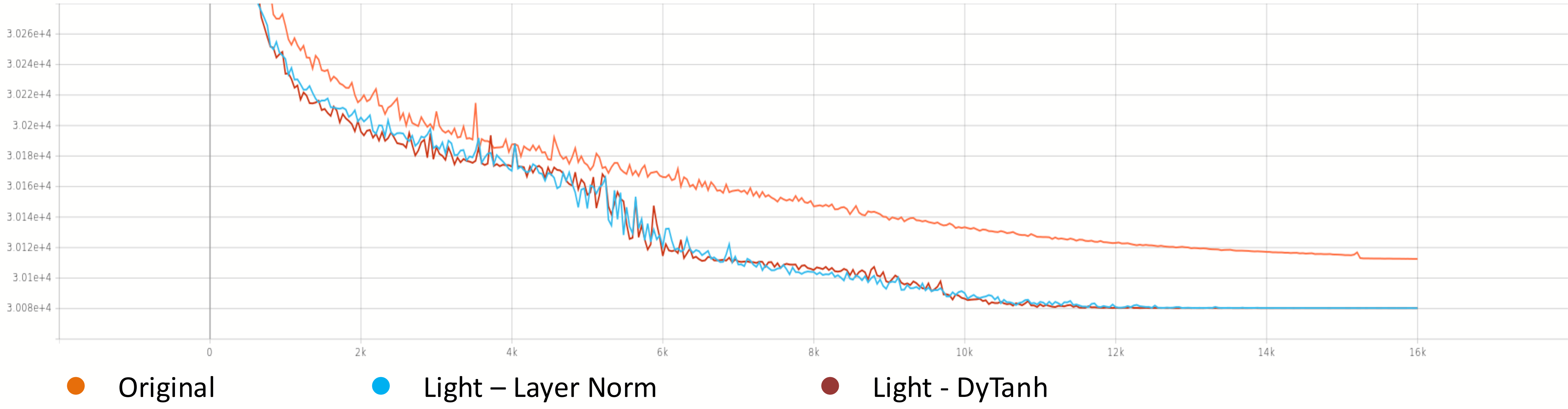
	Original	Light – Layer Norm	Light - DyTanh
Pro-Score	0.3603	0.4472	0.4394
Training time (min.)	16.98	11.70	11.70
Inference time	23.74	23.00	23.05
GFLOPS	91.73	81.64	81.64
GMACS	128.93	126.24	126.24
# Parameters (M)	12.75	7.49	7.48



Results

BTAD

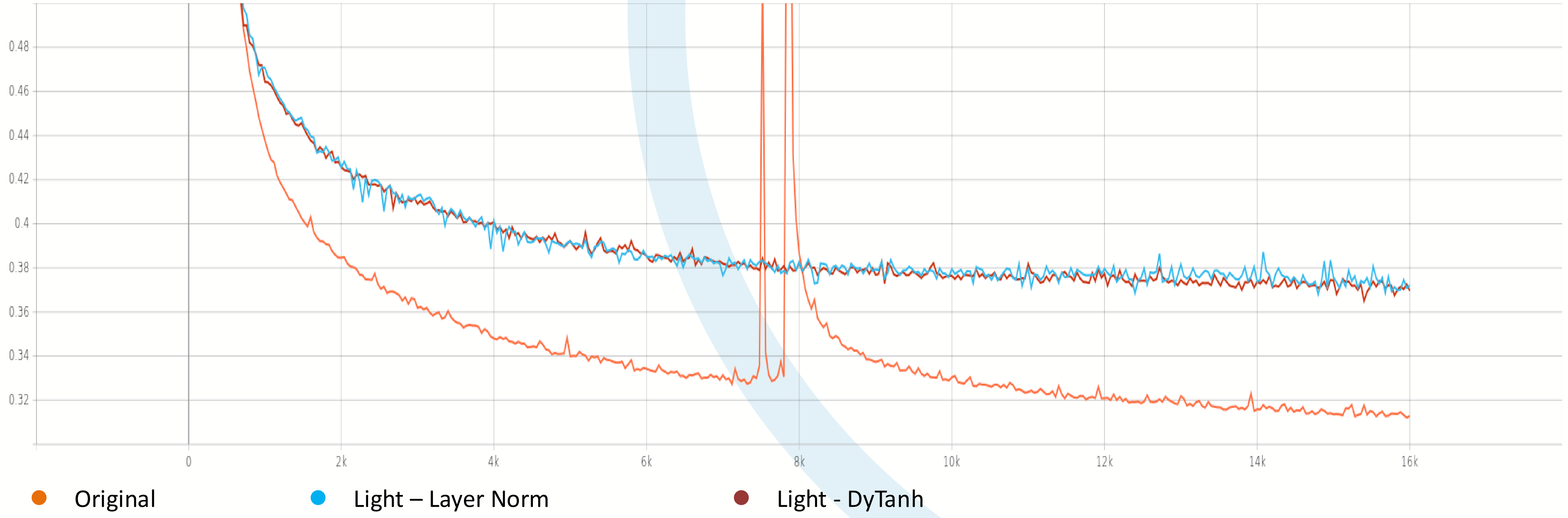
	Original	Light – Layer Norm	Light - DyTanh
Pro-Score	0.7007	0.6039	0.4369
Training time (min.)	31.24	20.76	20.92
Inference time	23.76	23.00	23.06
GFLOPS	91.73	82.73	82.73
GMACS	128.93	126.78	126.77
# Parameters (M)	12.75	8.54	8.53



What Happened?

The comulative loss is not everything:

- The original implementation has higher loss
- But lower SSIM loss!



Conclusions

- Hyperparameter tuning for efficiency matters
- Dynamic Tanh didn't show notable results
  - Maybe in future (optimized) implementations

1
2
3
4
5
6
7
8
9
10
11
12
13

References

- Mishra P., Verk R., Fornasier D., Piciarelli C., Foresti G. L., *VT-ADL: A Vision Transformer Network for Image Anomaly Detection and Localization* (2021)
- Zhu, J., Chen, X., He, K., LeCun, Y., & Liu, Z., *Transformers without Normalization* (2025)
- Liu J., Xie G., Wang J., Li S., Wang C., Zheng F., Jin Y., *Deep Industrial Image Anomaly Detection: A Survey* (2024)
- Bergmann P., Fauser M., Sattlegger D., Steger C., *MVTec AD — A Comprehensive Real-World Dataset for Unsupervised Anomaly Detection* (2021)
- Sabour S., Frosst N., Hinton G. E., *Dynamic Routing Between Capsules* (2017)

Computer Vision
2024/2025

Valerio Massimo Carioti

Thanks for your time