Supplementary Material: Weakly-Supervised Semantic Segmentation via Transformer Explainability

Anonymous Author(s)

The code of our reproducability attempt can be found at https://anonymous.4open.science/r/ViT_Affinity_Reproducibility_Challenge-7FBC

Qualitative Results on ImageNet - ViT Explainability (1)

In here, we provide qualitative results of the reproduced ViT explainability approach as proposed in (1)

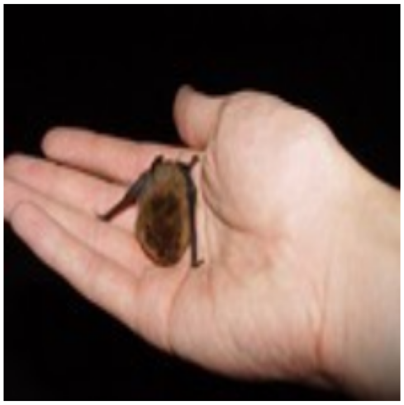

Figure 1: Image of a bug from ImageNet segmentation dataset (2).

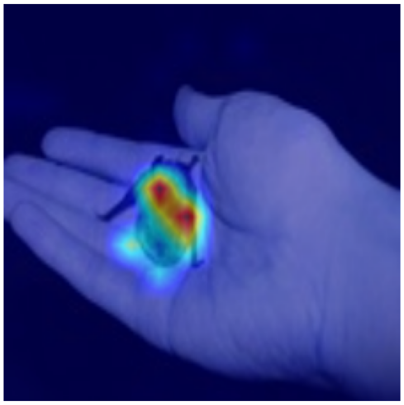

Figure 2: Segmentation map generated by our ViT-base for the bug image.

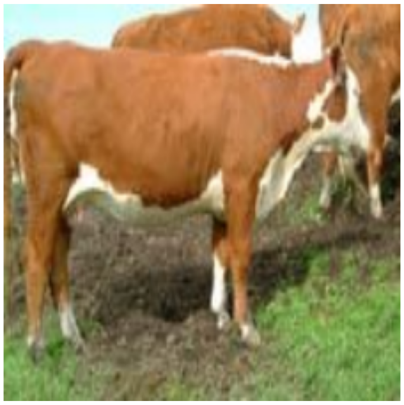

Figure 3: Image of a cow from ImageNet segmentation dataset (2).

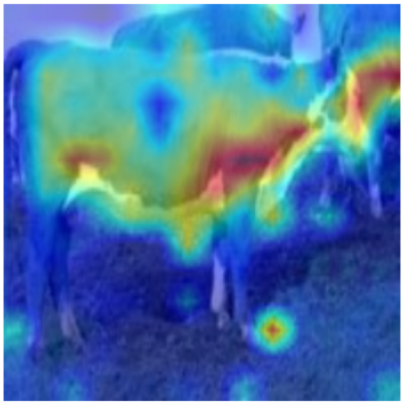

Figure 4: Segmentation map generated by our ViT-base for the cow image.

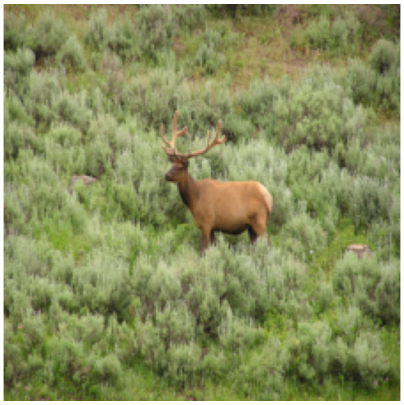

Figure 5: Image of a reindeer from ImageNet segmentation dataset (2).

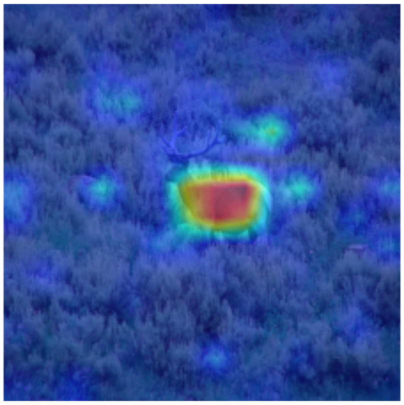

Figure 6: Segmentation map generated by our ViT-base for the reindeer image.

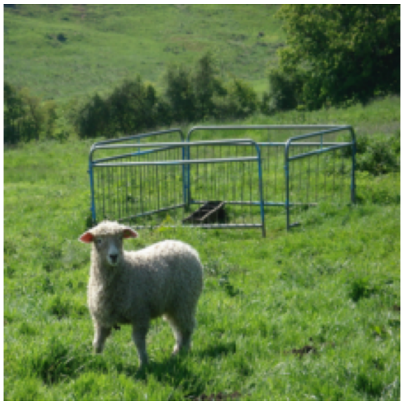

Figure 7: Image of a sheep from ImageNet segmentation dataset (2).

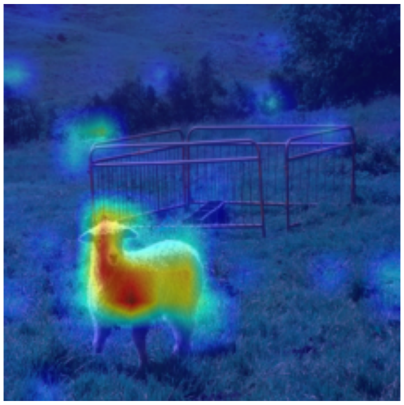

Figure 8: Segmentation map generated by our ViT-base for the sheep image.

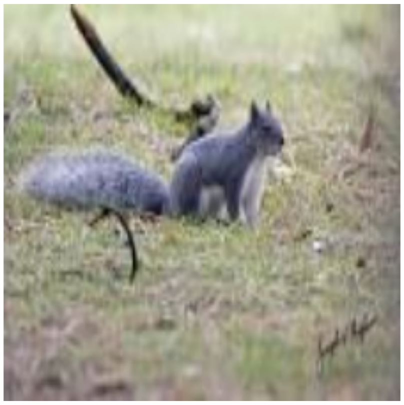

Figure 9: Image of a squirrel from ImageNet segmentation dataset (2).

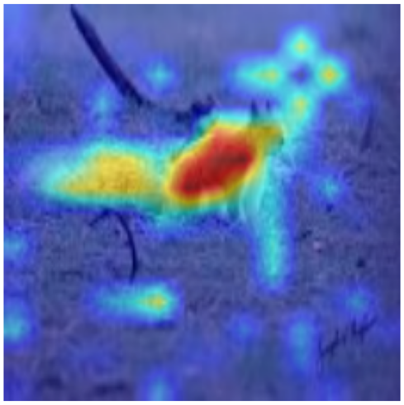

Figure 10: Segmentation map generated by our ViT-base for the squirrel image.

Qualitative Results on Pascal VOC - AffinityNet on Hybrid ViT

In here, we provide qualitative results of the reproduced ViT explainability approach as proposed in (1)

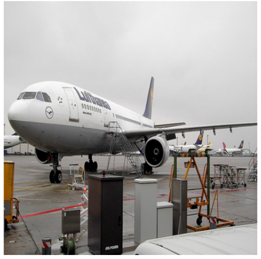

Figure 11: Image of an airplane from Pascal VOC segmentation dataset (3).

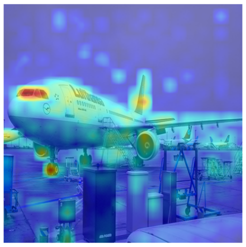

Figure 12: Segmentation map generated by our ViT-base for the airplane image.

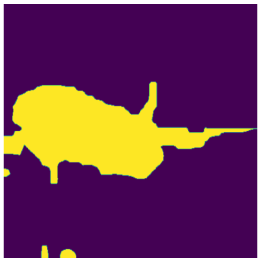

Figure 13: Affinity map generated by our AffinityNet for the airplane image.

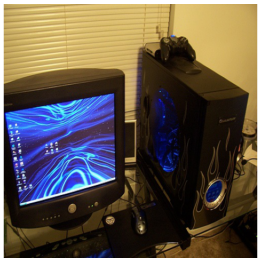

Figure 14: Image of an screen from Pascal VOC segmentation dataset (3).

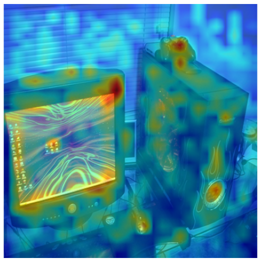

Figure 15: Segmentation map generated by our ViT-base for the screen image.

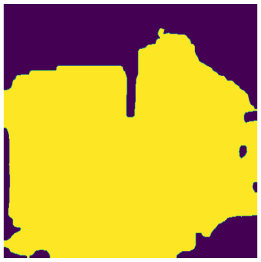

Figure 16: Affinity map generated by our AffinityNet for the screen image.

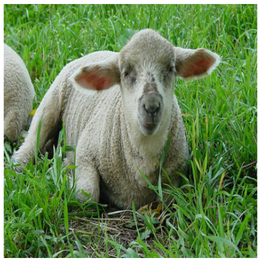

Figure 17: Image of a sheep from Pascal VOC segmentation dataset (3).

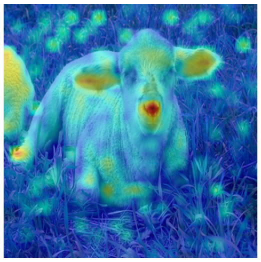

Figure 18: Segmentation map generated by our ViT-base for the sheep image.

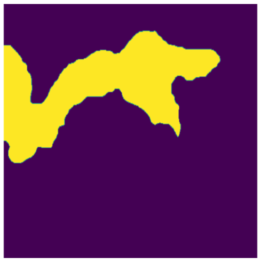

Figure 19: Affinity map generated by our AffinityNet for the sheep image.

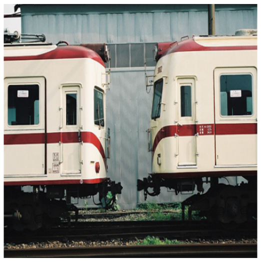

Figure 20: Image of a train from Pascal VOC segmentation dataset (3).

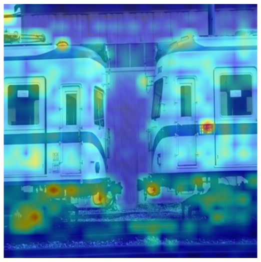

Figure 21: Segmentation map generated by our ViT-base for the train image.

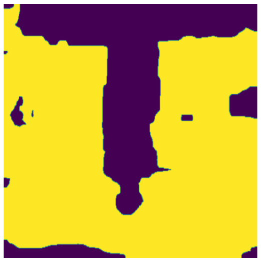

Figure 22: Affinity map generated by our AffinityNet for the train image.

References

[1] H. Chefer, S. Gur, and L. Wolf, "Transformer interpretability beyond attention visualization," *CoRR*, vol. abs/2012.09838, 2020.

[2] M. Guillaumin, D. Küttel, and V. Ferrari, "Imagenet auto-annotation with segmentation propagation," *International Journal of Computer Vision*, vol. 110, pp. 328–348, Dec. 2014.

[3] J. Ahn and S. Kwak, "Learning pixel-level semantic affinity with image-level supervision for weakly supervised semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4981–4990, 2018.