000
001
002
003
004
005
006
007
008
009
010
011
012
013
014
015
016
017
018
019
020
021
022
023
024
025
026
027
028
029
030
031
032
033
034
035
036
037
038
039
040
041
042
043
044
045
046
047
048
049
050
051
052
053

054
055
056
057
058
059
060
061
062
063
064
065
066
067
068
069
070
071
072
073
074
075
076
077
078
079
080
081
082
083
084
085
086
087
088
089
090
091
092
093
094
095
096
097
098
099
100
101
102
103
104
105
106
107

USAGE: A Unified Seed Area Generation Paradigm for Weakly Supervised Semantic Segmentation
-Supplementary Material-

Anonymous ICCV submission

Paper ID 1340

In this supplementary material, we provide more segmentation results (as shown in Fig. 1) as well as seed area comparison among different variants (as shown in Fig. 2).
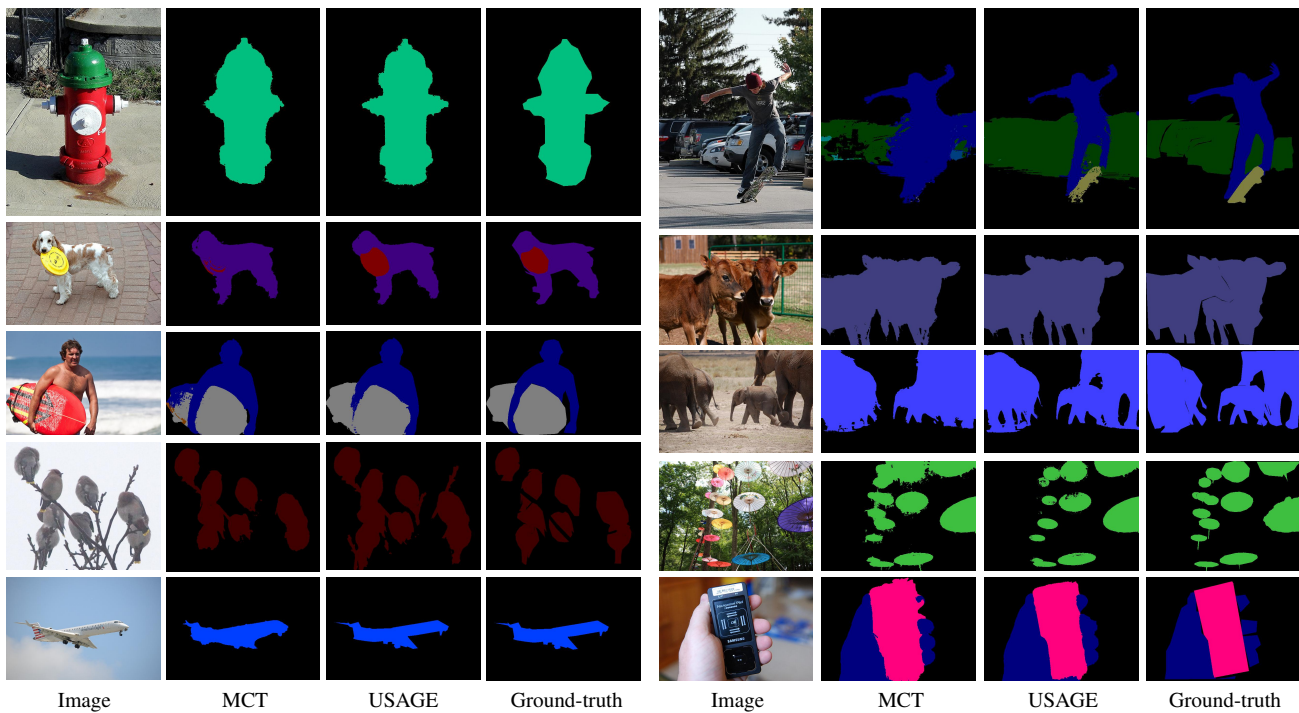
1. More Segmentation Results

| Image | MCT | USAGE | Ground-truth | Image | MCT | USAGE | Ground-truth |

Figure 1. Segmentation results on the MS COCO [2] *val* set. MCT: Class-to-patch attention map [6].

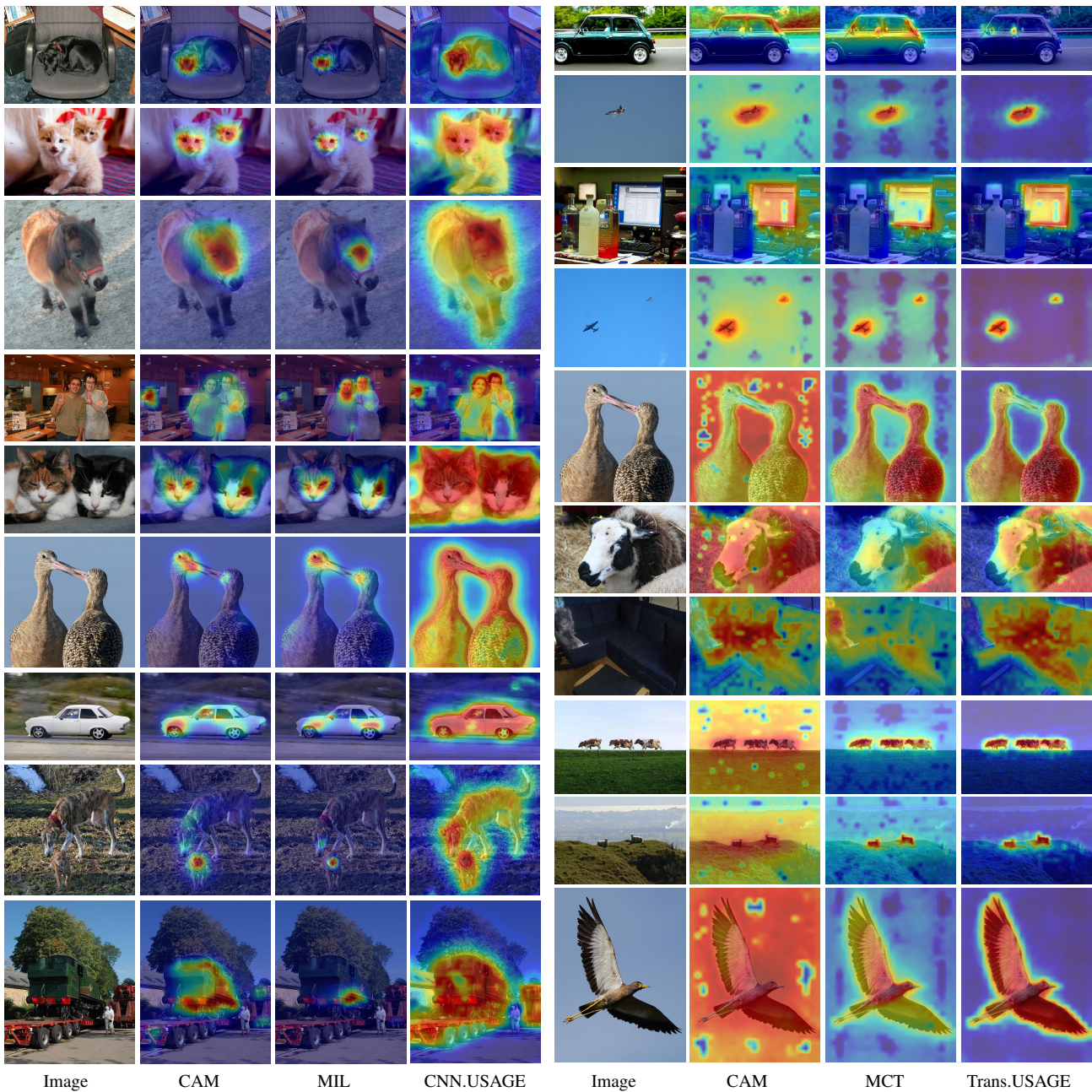## 2. Seed Area Comparison among Different Variants

Figure 2. Seed area visualizations on the PASCAL VOC [1] *train* set. Left: Seed area generation from a CNN (ResNet38 [5]). Right: Seed area generation from a Transformer (DeiT-S [4]). MCT: Class-to-patch attention map [6]. MIL: a MIL-based method [3].

References

[1] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *IJCV*, 88(2):303–338, 2010. 2

[2] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *ECCV*, 2014. 1

[3] Pedro O Pinheiro and Ronan Collobert. From image-level to pixel-level labeling with convolutional networks. In *CVPR*, 2015. 2

ICCV
#1340

ICCV
#1340

ICCV 2023 Submission #1340. CONFIDENTIAL REVIEW COPY. DO NOT DISTRIBUTE.

[4] Hugo Touvron, Matthieu Cord, Matthijs Douze, Francisco Massa, Alexandre Sablayrolles, and Hervé Jégou. Training data-efficient image transformers & distillation through attention. In *ICML*, 2021. 2

[5] Zifeng Wu, Chunhua Shen, and Anton Van Den Hengel. Wider or deeper: Revisiting the resnet model for visual recognition. *Pattern Recognition*, 90:119–133, 2019. 2

[6] Lian Xu, Wanli Ouyang, Mohammed Bennamoun, Farid Boussaid, and Dan Xu. Multi-class token transformer for weakly supervised semantic segmentation. In *CVPR*, 2022. 1, 2