Enhancing Material Features Using Dynamic Backward Attention on Cross-Resolution Patches

Yuwen Heng, Yihong Wu, Srinandan Dasmahapatra, Hansung Kim
University of Southampton, UK
{y.heng, yihongwu}@soton.ac.uk, sd@ecs.soton.ac.uk, h.kim@soton.ac.uk

Introduction

- Applications: robotic system, vision-based acoustic and optical property estimation.
- Challenge of image-based material segmentation:
  - A specific material can have a variety of appearances, such as shape, colour and transparency.

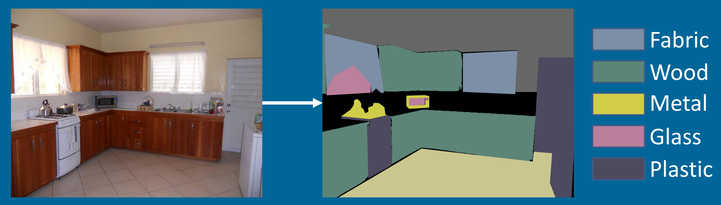


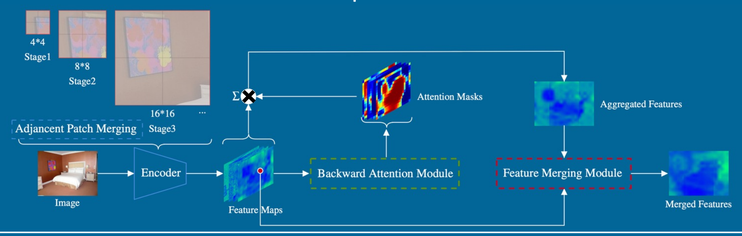
Image Segmented materials

Methodology

- Idea: combining material features and contextual features.
- Material features (in image patches) allow the network to identify the categories without covering all varied appearances.
- Contextual features (in full image) can limit the possible categories of materials that appear in a given scene.

Contextual features in full image Material features in image Patches

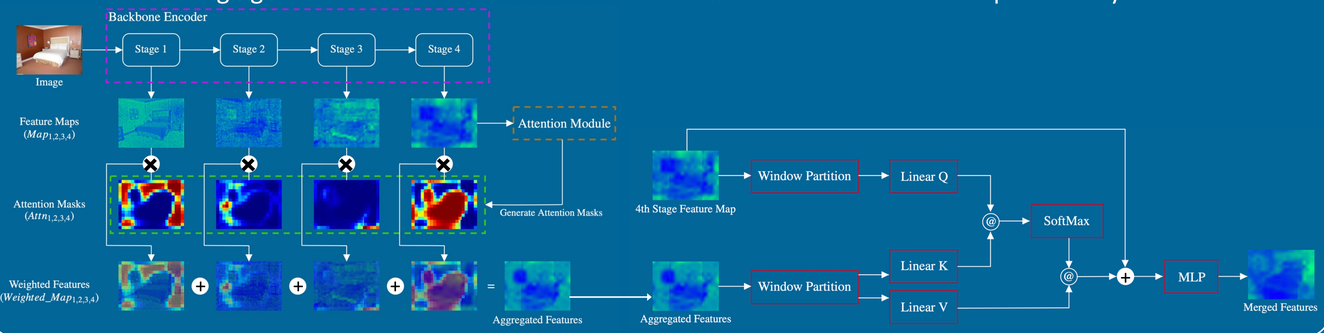
- Proposed method: segmenting the images with material features extracted from cross-resolution patches.



*ACK. This work was supported by the EPSRC Programme Grant Immersive Audio-Visual 3D Scene Reproduction Using a Single 360 Camera (EP/V03538X/1).

Network Architecture

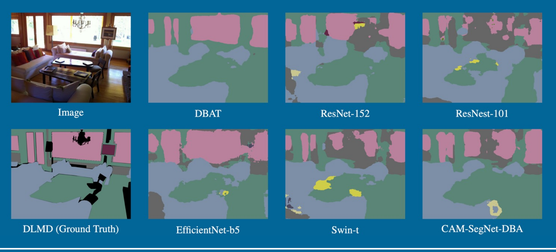
- The encoder backbone provides cross-resolution features by merging adjacent patches at each transformer stage.
- The backward attention module aggregates these cross-resolution features.
- The feature merging module with a residual connection makes the network learn complementary features.



Analysis

- The DBAT outperforms the second-best model in this paper by 2.15% in pixel accuracy.
- The boundary quality is more adequate than the segments predicted by other networks.
- All chosen models can work in real-time.

Datasets / Architecture	LMD Pixel Acc	LMD Mean Acc	OpenSurfaces Pixel Acc	OpenSurfaces Mean Acc	OpenSurfaces mIoU	– #params (M)	– #flops (G)	– FPS
ResNet-152	80.68 ± 0.11	73.87 ± 0.25	83.80	63.56	52.09	60.75	70.27	31.35
ResNeSt-101	82.45 ± 0.20	75.31 ± 0.29	85.10	67.13	55.32	48.84	63.39	25.57
EfficientNet-b5	83.17 ± 0.06	76.91 ± 0.06	84.63	65.47	53.25	30.17	20.5	27.00
Swin-t	84.70 ± 0.26	79.06 ± 0.46	86.19	69.41	57.71	29.52	34.25	33.94
CAM-SegNet-DBA	86.12 ± 0.15	79.85 ± 0.28	**86.64**	69.92	**58.18**	68.58	60.83	17.79
DBAT	**86.85 ± 0.08**	**81.05 ± 0.28**	86.28	**70.68**	58.08	56.03	41.23	27.44



Conclusion & Future Work

- Our DBAT beats all chosen models that can serve real-time applications on two datasets, and achieves comparable performance with fewer FLOPs than the multi-branch CAM-SegNet [2].
- In the future, we plan to interpret the material features that our DBAT learns by comparing them with features extracted from different tasks, such as object segmentation.

References

- [1] Gabriel Schwartz and Ko Nishino. Recognizing material properties from images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 42(8):1981–1995, 2020. doi: 10.1109/TPAMI.2019.2907850.
- [2] Yuwen Heng, Yihong Wu, Hansung Kim, and Srinandan Dasmahapatra. Cam-segnet: A context-aware dense material segmentation network for sparsely labelled datasets. In 17th International Conference on Computer Vision Theory and Applications (VISAPP), volume 5, pages 190–201, 2022.