

Learning Semantic Neural Tree for Human Parsing Supplementary Material

Ruyi Ji^{1,2}, Dawei Du³, Libo Zhang^{1, *}, Longyin Wen⁴,
Yanjun Wu¹, Chen Zhao¹, Feiyue Huang⁵, and Siwei Lyu³

¹ Institute of Software Chinese Academy of Sciences, China,

² University of Chinese Academy of Sciences, China,

³ University at Albany, State University of New York, USA,

⁴ JD Finance America Corporation, USA,

⁵ Tencent Youtu Lab, China.

1 The Category Label Definition in the LIP and CIHP Datasets

Since the LIP [4] and CIHP [3] datasets use the same annotations, we adopt the same architecture of our neural tree, shown in Figure 1. We report the category label definition in our neural tree in Table 1.

Table 1. The category label definition used in the LIP [4] and CIHP [3] datasets.

Leaf	Label
L_1^1	Background
L_1^2	Hat, Hair, Sunglasses, Face
L_1^3	Scarf, Upper-clothes, Coat, Dress
L_2^3	Left-arm, Right-arm, Glove
L_3^3	Skirt, Pants, Jumpsuit, Left-leg, Right-leg
L_4^3	Socks, Left-shoe, Right-shoe

2 The Architecture of Our Semantic Neural Tree in the Pascal-Person-Part Dataset

The PASCAL-Person-Part dataset [1] is originally from the PASCAL VOC-2010 dataset [2], and is extended for human parsing with 6 coarse body part labels (*i.e.*, *head*, *torso*, *upper-/lower-arms*, and *upper-/lower-legs*). As shown in Figure 2, we construct a neural tree with 3-level. We summarize the category label definition in Table 2.

* Libo Zhang is the corresponding author(libo@iscas.ac.cn).

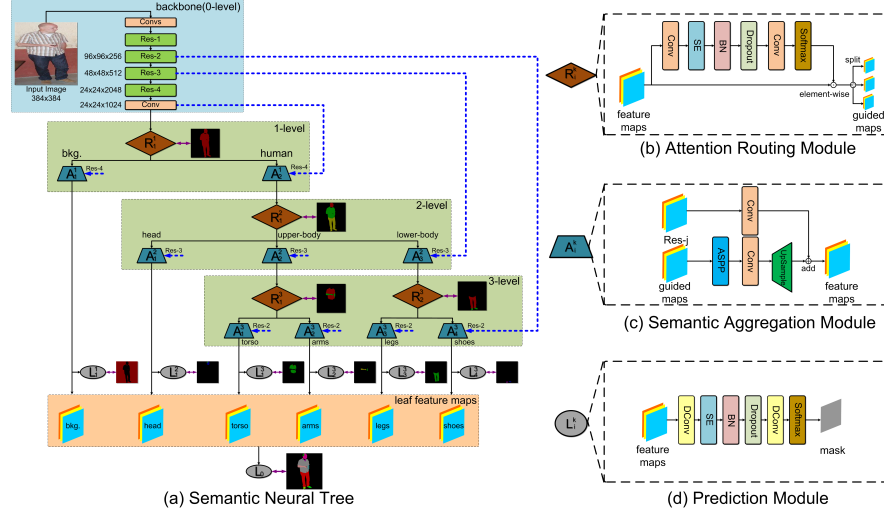


Fig. 1. The architecture of our semantic neural tree used in the LIP [4] and CIHP [3] datasets.

Table 2. The category label definition used in the Pascal-Person-Part dataset.

Leaf	Label
L_1^1	Background
L_1^2	Head
L_1^3	Torso
L_2^3	U-arms
L_3^3	L-arms
L_4^3	U-legs
L_5^3	L-legs

3 The Architecture of Our Semantic Neural Tree in the MHP-v2 Dataset

The MHP-v2 dataset [5] includes 25,403 elaborately annotated images with 58 fine-grained semantic category labels, involving 2 ~ 26 persons per image and captured in real-world scenes from various viewpoints, poses, occlusion, interactions and background. As shown in Figure 3, we construct a neural tree with 5-level. We summarize the category label definition in Table 3.

References

1. Chen, X., Mottaghi, R., Liu, X., Fidler, S., Urtasun, R., Yuille, A.L.: Detect what you can: Detecting and representing objects using holistic models and body parts. In: CVPR. pp. 1979–1986 (2014)

2. Everingham, M., Gool, L.V., Williams, C.K.I., Winn, J.M., Zisserman, A.: The pascal visual object classes (VOC) challenge. *IJCV* **88**(2), 303–338 (2010)
3. Gong, K., Liang, X., Li, Y., Chen, Y., Yang, M., Lin, L.: Instance-level human parsing via part grouping network. In: *ECCV*. pp. 805–822 (2018)
4. Liang, X., Gong, K., Shen, X., Lin, L.: Look into person: Joint body parsing & pose estimation network and a new benchmark. *TPAMI* **41**(4), 871–885 (2019)
5. Zhao, J., Li, J., Cheng, Y., Sim, T., Yan, S., Feng, J.: Understanding humans in crowded scenes: Deep nested adversarial learning and A new benchmark for multi-human parsing. In: *ACM MM*. pp. 792–800 (2018)

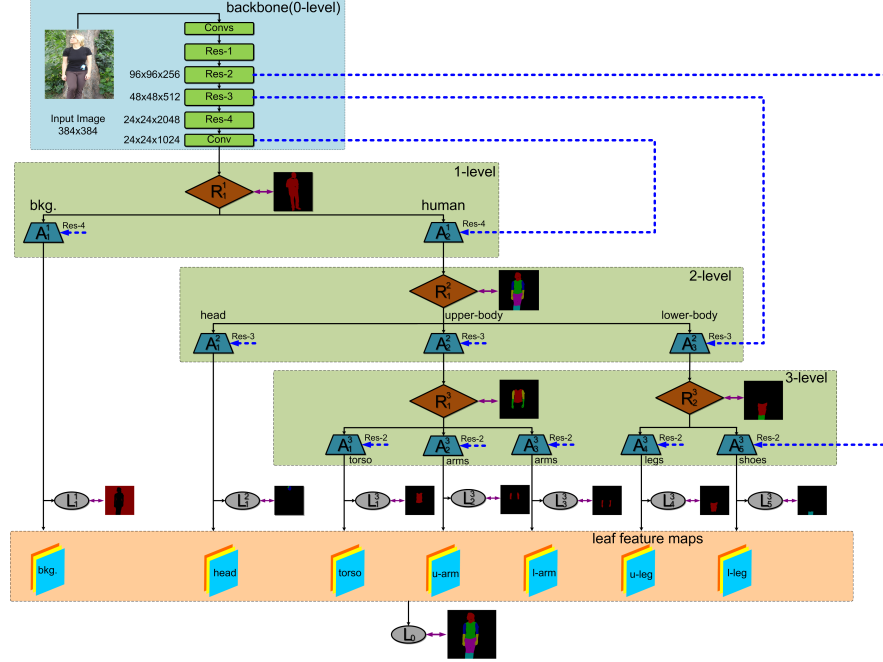


Fig. 2. The architecture of our semantic neural tree used in the Pascal-Person-Part dataset [1].

Table 3. The category label definition used in the MHP-v2 dataset.

Leaf	Label
L_1^1	Background
L_1^2	Backpack, Protector, Ball, Bats, Bottle, Carrybag, Cases, Umbrella, Wallet/Purse
L_1^3	Other-full-body-clothes, Other-accessary, Other-upper-body-clothes, Other-lower-body-clothes
L_1^4	Cap/Hat, Helmet, Hair, Sunglasses, Face, Headwear, Eyewear
L_1^5	Bikini/Bra, Jacket/Windbreaker/Hoodie, Tee-shirt, Polo-shirt, Sweater, Singlet, Torso-skin, Robe, Coat, Dress, Tie, Scarf, Belt
L_2^5	Glove, Watch, Wristband, Left-arm, Right-arm, Left-hand, Right-hand
L_3^5	Left-leg, Right-leg, Jumpsuit, Pants, Shorts/Swim-shorts, Skirt
L_4^5	Stockings, Socks, Left-boot, Right-boot, Left-shoe, Right-shoe, Left-highheel, Right-highheel, Left-sandal, Right-sandal, Left-foot, Right-foot

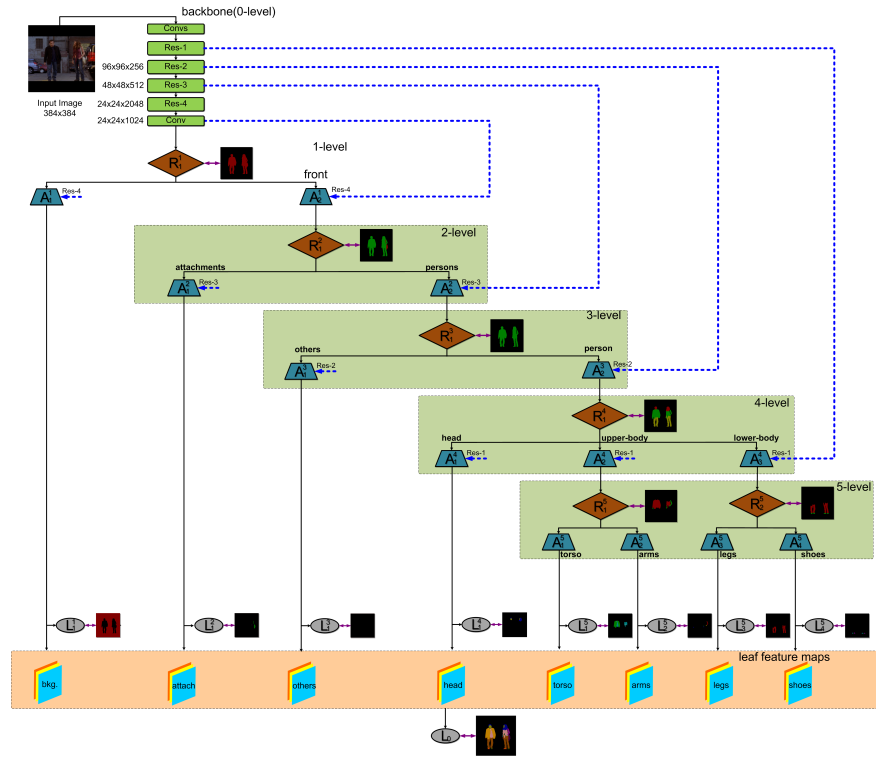


Fig. 3. he architecture of our semantic neural tree used in the MHP-v2 dataset [5].