# Supplementary Materials for: Border-ownership, Category-selectivity and Beyond

*Dataset Preparation and Results*

Tianlong Chen[1], Xuemei Cheng[2] and Thomas Tsao[3]
[1,2,3]OpticArray Technologies, Inc. 1917 Dundee Rd., Rockville MD 20850 USA

{tc, xmcheng, ttsao}@opticarraytech.com

We have selected a few varieties of different datasets, none of them was originally designed for the purpose for the border ownership assignments. We use ground truth sets re-generated from original datasets to train and test TcNet.

## 1. FlyingChair Dataset

For the FlyingChairs dataset, we selected 1390 different chair-types from the original FlyingChairs dataset [19], each of which had 62 different orientations. We used 38 different images randomly downloaded from the internet for image backgrounds. Each image of the training set (resolution 512x1024) was generated by randomly selecting a portion of one of the randomly-selected background images and adding a random number (between 1-8) of randomly-selected chair types with random orientation at random positions. The order of the chairs was recorded, allowing accurate determination of border ownership**.** Ground truth contour maps and border-ownership maps were generated accordingly (Fig. S1). We generated 10000 images for the training set, 1000 images for the validation set, and 1000 images for the test set. The dataset has one category of 'chairs', so no category channel is included.

Fig.S2 demonstrates a sample inferred result from trained TcNet and its ground truth as comparison. A 2-branch TcNet was trained on the dataset. The 1st row in Table 1 shows the inference evaluation statistics for FlyingChair dataset.

## 2. Synscapes Dataset

The Synscapes dataset from [17] is a computer-generated synthetic street scene dataset; we used the first 15000 images from the original dataset (original resolution 720x1440) for training, each of which was split into left and right halves such that the total number of images in the training set was 30000 (resolution 704x896). We then selected 1000 separate images from the original dataset for validation, center-cropping each to the same resolution (704x896). For testing, we used the remaining 4000 images from the original dataset, center-cropping each to a resolution of 704x1408. We defined three categories: 'person' (including people on bicycles), 'vehicle' (including cars, buses, and trains), and 'traffic sign and block'. The depth information from the original dataset was used to determine the occluding order of overlapping objects, i.e., border ownership.

Fig. S3 demonstrates the Synscapes dataset and sample ground truth and inference results of border-ownership coding together with 3 categories. A 5-branch TcNet with 3 category channels was trained on the re-generated dataset; 5-branch (1+1+3) being 1 enforcement branch, 1 border-ownership branch, and 3 category branches. The 2$^{nd}$ row in Table 1 shows the inference evaluation statistics for Synscapes dataset including border-ownership and 3 selected category channels.

## 3. Plant Dataset

The plant dataset was from [18]; it has one category of 'leaf', so no category-channel included. The dataset was chosen to see how occluding contour and border-ownership coding works for conceptually separated leaves of a plant ('conceptual' or 'virtual' does not actually or physically exist but exist virtually in brain imagination as visual concept); and it is worth mentioning that an interesting dataset expansion method was used on this small footprint original dataset. For the Plant dataset, we selected the Arabidopsis set from the Plant Phenotyping Dataset [18] (128 original ground truth including image and leaf instance label, but no depth or other clues for occluding orders). Due to the limited number of images, we generated a training set using the leaves from the 128 images (1938 leaves in total). For each generated image, the background was cropped from a few images from the original dataset and tile-packed, and a random number of leaves (between 4-15) was randomly positioned around the center of the image. The order of the leaves was recorded, allowing accurate determination of border ownership. Examples of generated data are shown in Fig. S4. We generated 5000 images for the training set and 500 images for the validation set.

Fig. S5. shows the inferred results of two samples after 2-branch TcNet was trained with the random-leaf training set. *It was observed that the leaves were not separated conceptually, as no top-down guidance exists, suggesting that conceptual or virtual boundary with no fixed expectation (leaves were part of plant not separately 'expected') may need active top-down guidance for conceptual separation*.

## 4. **Face Dataset**

We selected the Face dataset because face is one of limited number of category-selective targets in human or mammal vision system [23, 24]; knowing the behavior of TcNet on face dataset is of great interest. For the Face dataset, the original CelebAMask-HQ dataset [6] contained 30000 ground truth face images (resolution 512x512) with different face 'components' (eyes, eyebrows, lips, nose, and so on), which were split into train (25000), validation (2500) and test (2500) sets and resolution was kept the same. We defined 4 object categories: eyebrows, eyes, lips, and nose. the upper contour of a nose was removed as it is not a real object boundary.

Fig. S6. demonstrates samples of our re-generated Face dataset ground truth and inferred results from trained 6-branch TcNet, including border-ownership and 4 different category channels; The 3rd row of Table 1 shows the inference evaluation statistics of Face dataset. *It was observed that conceptual contours of blocked eyebrows or lips were detected in some cases, suggesting that conceptual or virtual boundary with fixed expectation (blocked eyebrows or lips was 'expected') may not need top-down guidance.*

**Fig. S1. Sample Training sets of 2-channel border ownership coding and contours for FlyingChair Dataset**. Each ground truth set includes an image (**top row**, which includes random number (between 1 to 8) of randomly selected chairs randomly positioned on a randomly selected background image), a 1-channel map of occluding contours (**middle row**) and a 2-channel map of border ownership coding (**bottom row**).

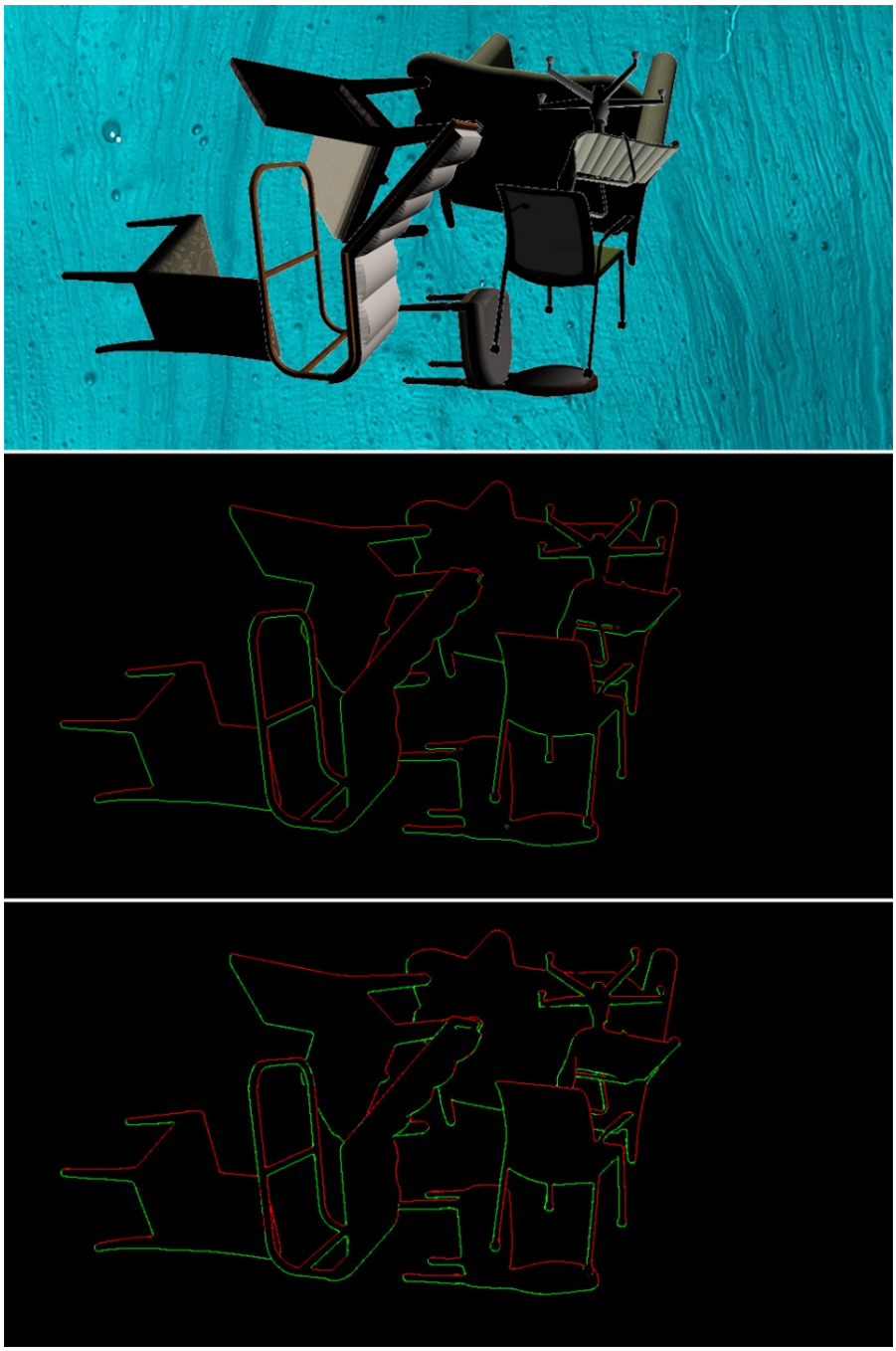

**Fig. S2. Example border ownership coding inference results for FlyingChair dataset**. **Top**: input testing image; **Middle**: ground truth border ownership map; **Bottom**: inferred border ownership map from the output of a trained 2-branch TcNet.

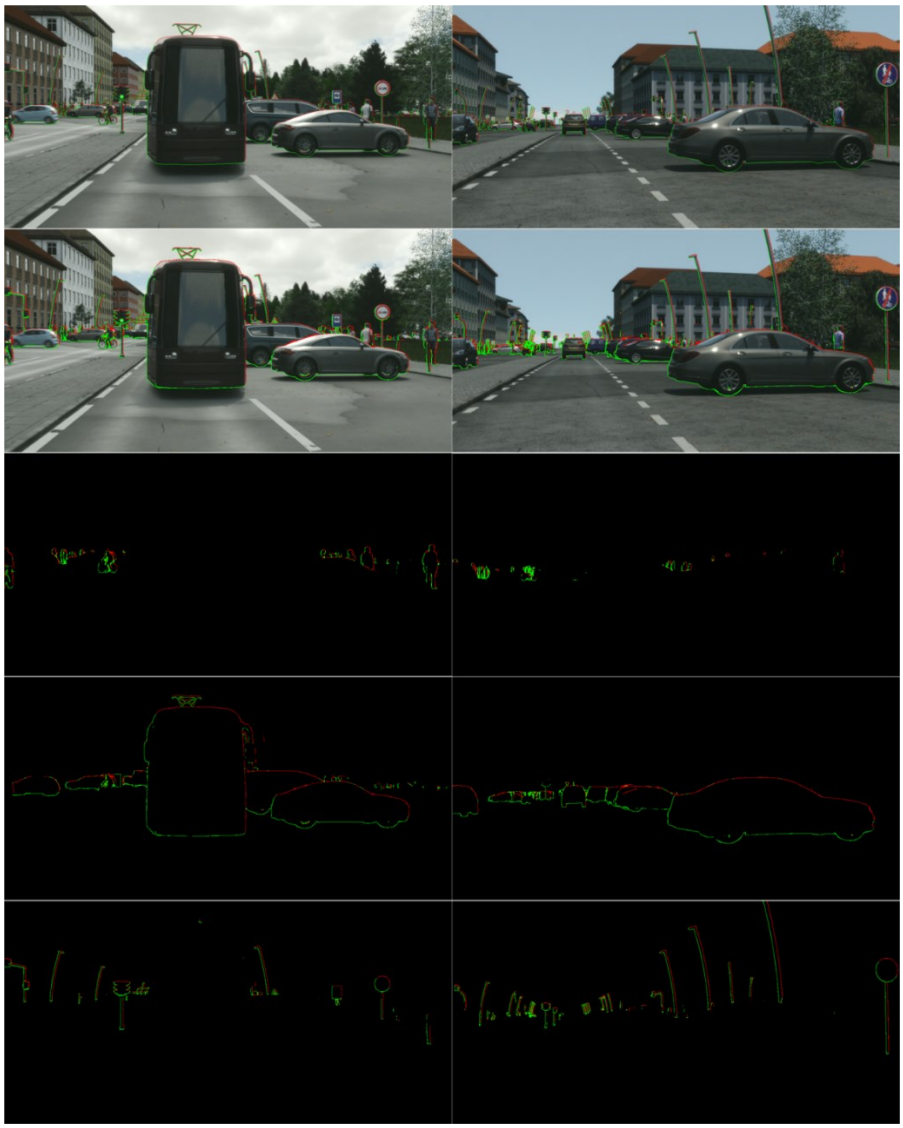

**Fig. S3 Example border ownership coding and category-selective inference results for Synscapes dataset**. **1st Row**: 2-channel border ownership ground truth for 3 categories ('person', 'vehicle' and 'traffic signs/blocks') overlayed over input images; **2nd Row**: inferred 2-channel and 3-category border-ownership map overlayed over input images; **3rd Row**: 'person' category border-ownership map; **4th Row**: 'vehicle' category border-ownership map; **5th Row**: 'traffic signs/blocks' category border-ownership map.

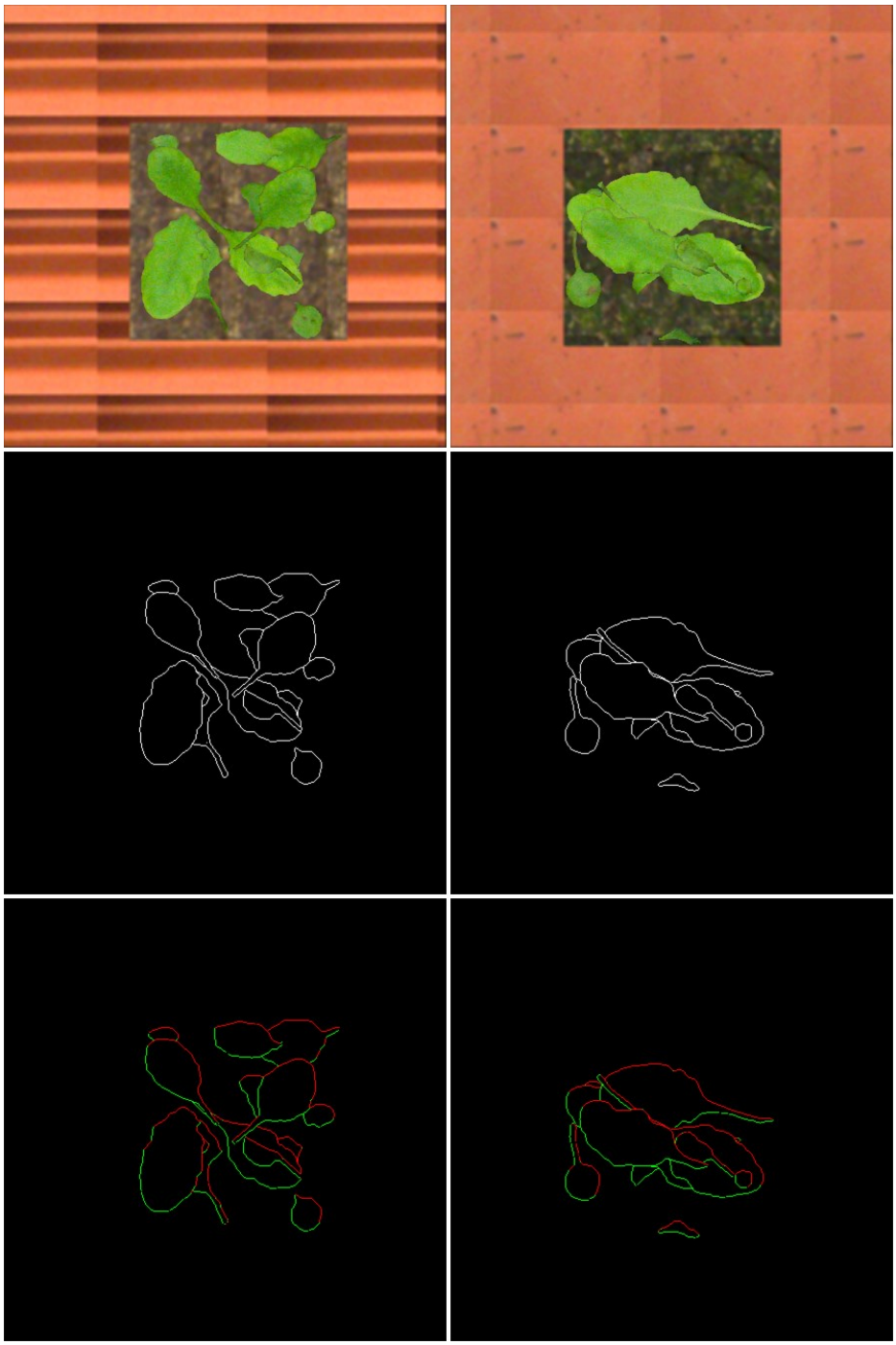

**Fig. S4. Example ground truth data for Plant dataset. Top row**: example generated images with a random number of leaves (between 4-15) randomly positioned around the center region of

images; **Middle row**: occluding contours of leaves; **Bottom row**: 2-channel border-ownership map of the leaves.

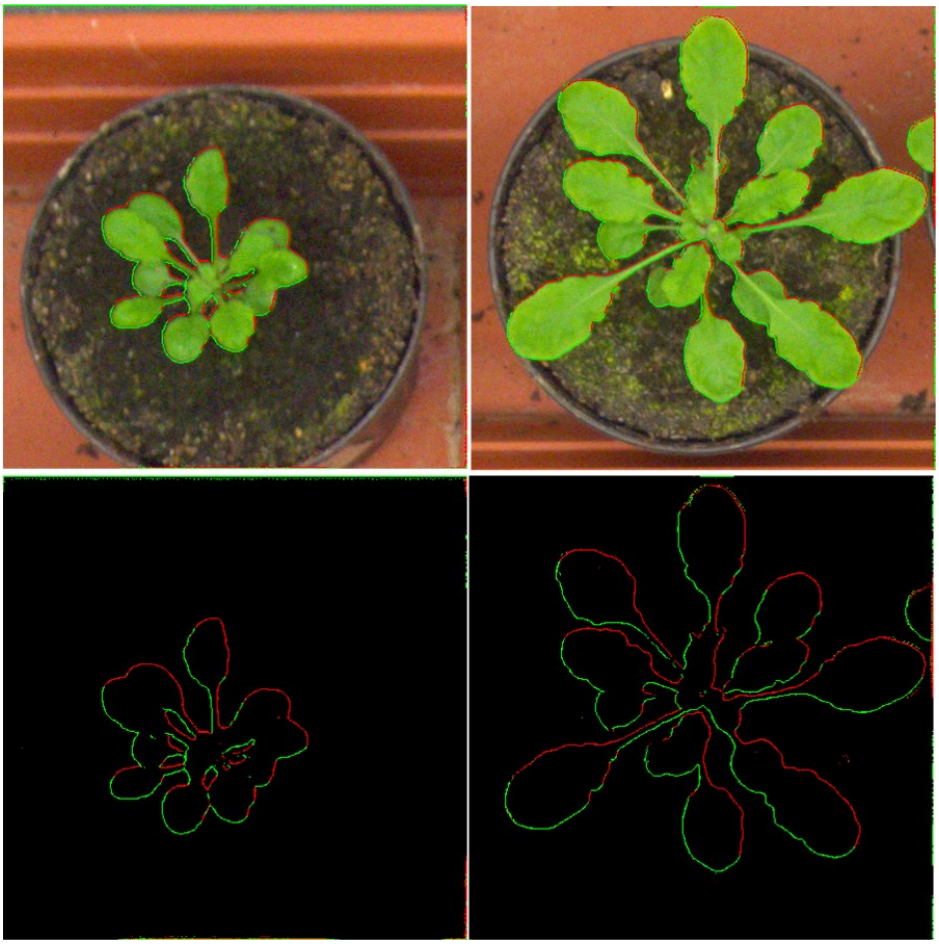

**Fig. S5. Example border ownership inference results for Plant dataset**. **Top row**: input image overlayed with 2-channel inferred border-ownership map; **Bottom row**: 2-channel inferred border-ownership map.

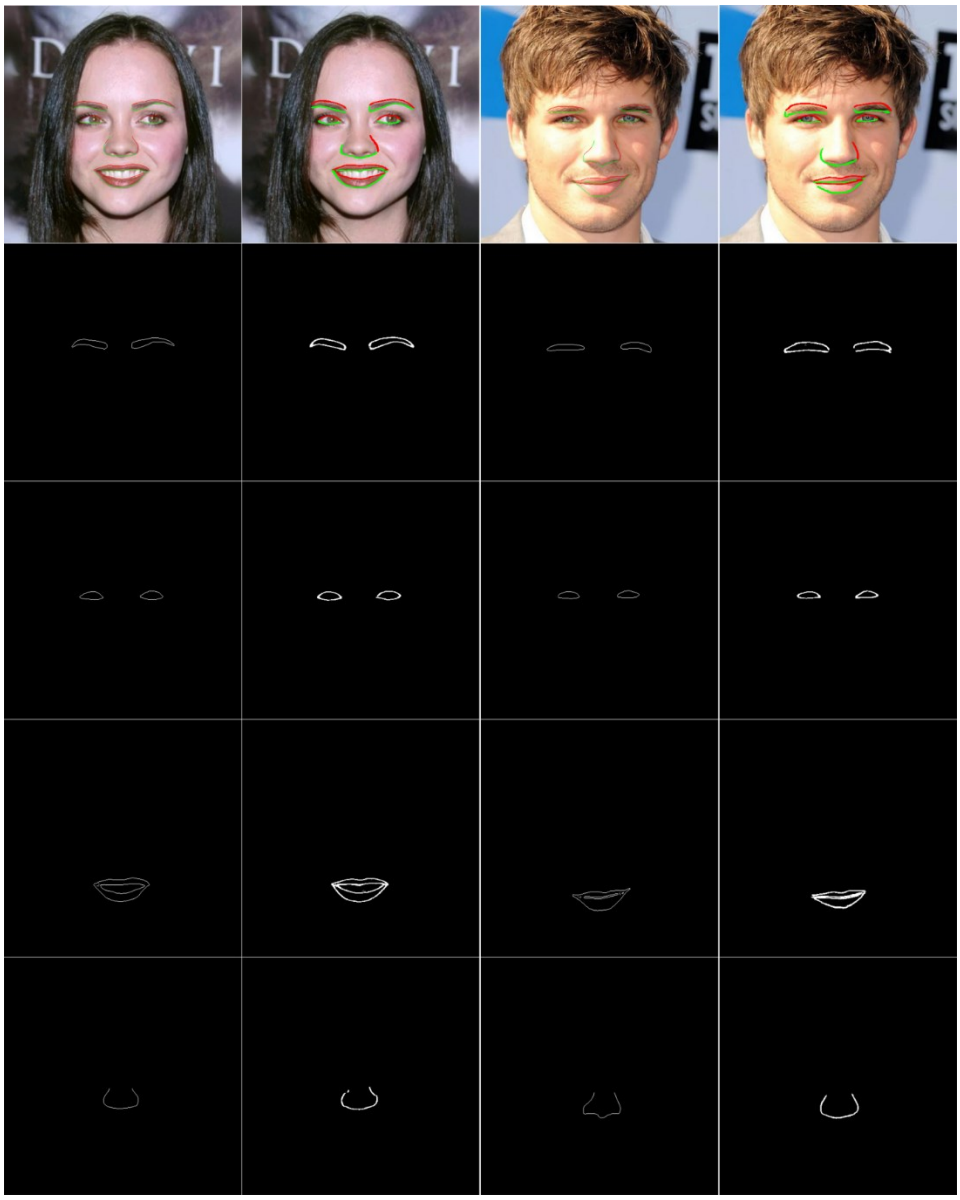

**Fig. S6. Example ground truth and inferred results of border-ownership and category maps for Face dataset**. Four categories are selected in the example ('eyebrow', 'eye', 'lips', and 'partial nose'). **1st and 3rd columns** are ground truth, and **2nd and 4th columns** are inferred results from the output of a trained 6-branch TcNet. **1st row**: 2-channel border-ownership map overlayed over images; **2nd row**: eyebrow channel; **3rd row**: eye channel; **4th row**: lip channel; **5th row**: nose channel.

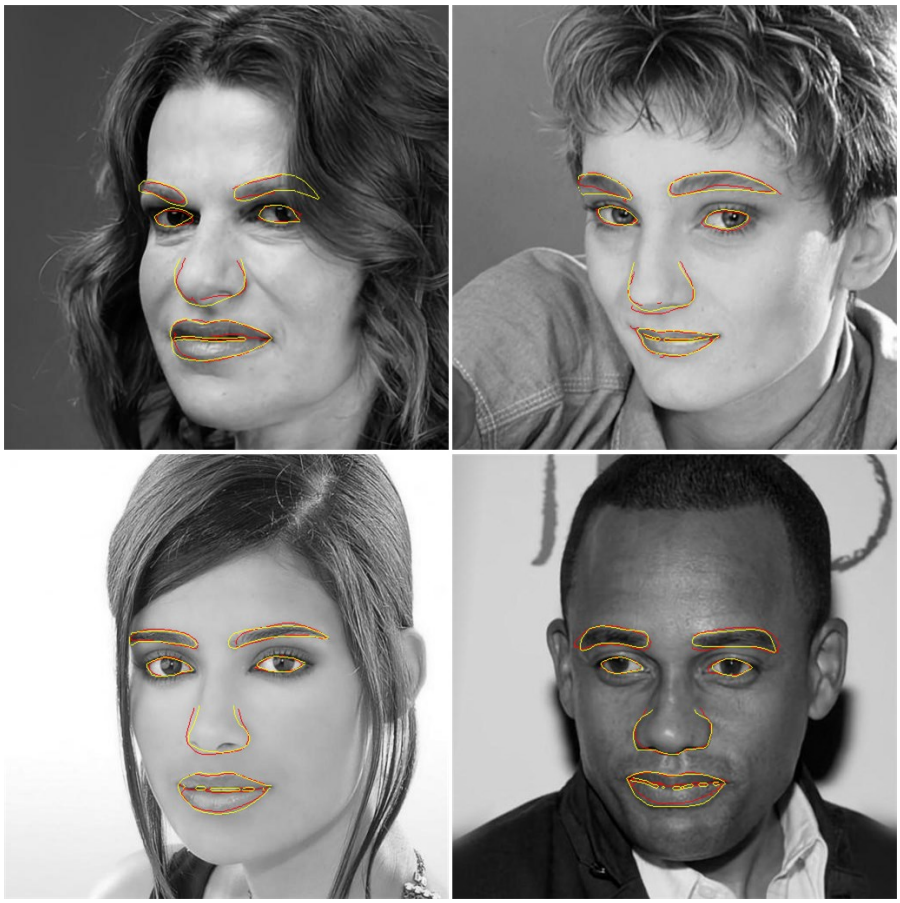

**Fig. S7. Examples of high-accuracy determination of object contours for Face dataset**. Red-colored inference contour and yellow-colored ground truth contour are overlayed over input images (grayscale version of original color images). Visual inspection suggests that inferred contours are more accurate than ground truth contours.

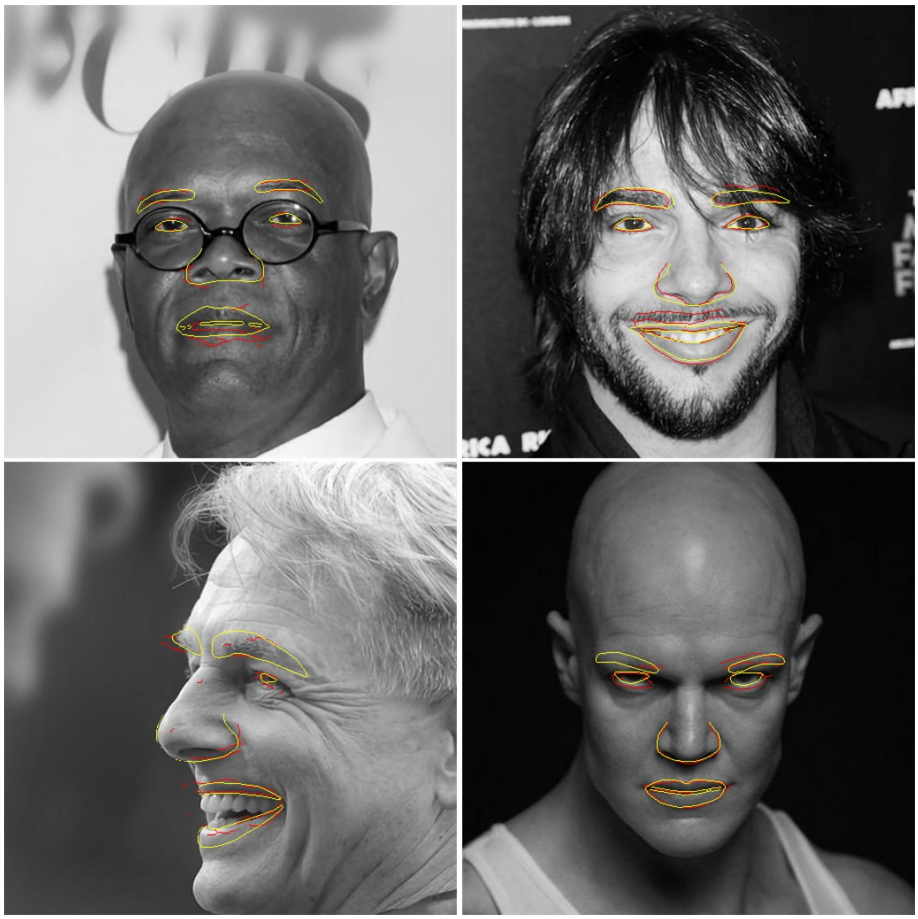

**Fig. S8. Example failure cases for Face dataset**. Red-colored inference contour and yellow-colored ground truth contour are overlayed over input images (grayscale version of original color images). Sample failure cases: profile face, no eyebrow, beard.

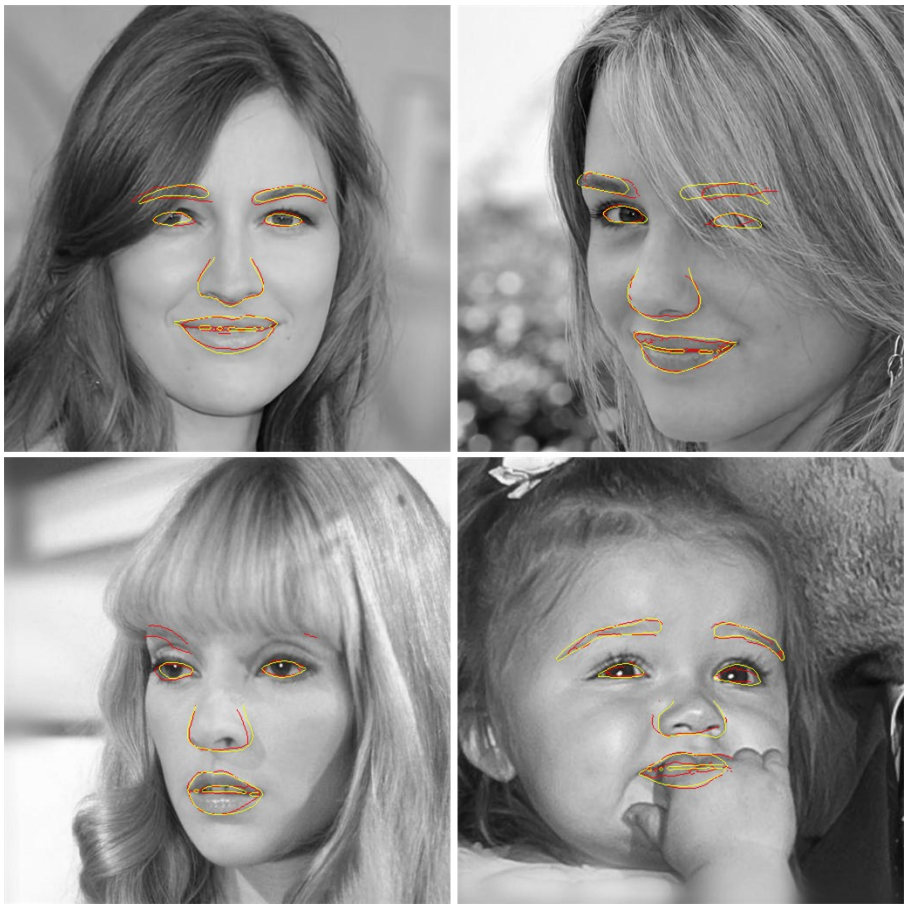

**Fig. S9. Examples of inferred conceptual contours.** Red-colored inference contour and yellow-colored ground truth contour are overlayed over input images (grayscale version of original color images). Some of the contours should be blocked by hands (**bottom right**) or hairs (**other three**).

## References


1. Alexander Kirillov, Kaiming He, Ross Girshick, Carsten Rother, and Piotr Dollár: Panoptic Segmentation. arXiv:1801.00868v3 [cs.CV] Apr. 10, 2019; https://www.cityscapes-dataset.com/
2. Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton: ImageNet Classification with Deep Convolutional Neural Networks. Communications of the ACM, Vol 60 No.6 doi:10.1145/3065386

3. Antoine Dedieu, Rajeev V. Rikhye, Miguel Lazaro-Gredilla, Dileep George: Learning attention-controllable border-ownership for objectness inference and binding. bioRxiv, https://doi.org/10.1101/2020.12.31.424926
4. Antonio Buades, Rafael Grompone von Gioi and Julia Navarro: Joint Contours, Corner and T-Junction Detection: An Approach Inspired by the Mammal Visual System. J Math Imaging Vis 2017 doi:10.1007/s10851-017-0763-z
5. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin: Attention is All You Need. arXiv:1706.03762v5 [cs.CL] Dec. 6, 2017
6. Cheng-Han Lee, Ziwei Liu, Lingyun Wu, and Ping Luo: MaskGAN: Towards Diverse and Interactive Facial Image Manipulation. arXiv:1907.11922 [cs.CV] Apr. 1, 2020; https://github.com/switchablenorms/CelebAMask-HQ
7. Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich: Going deeper with convolutions. arXiv:1409.4842v1 [cs.CV]
8. D. George, W. Lehrach, K. Kansky, M. Lázaro-Gredilla, C. Laan, B. Marthi, X. Lou,, Z. Meng, Y. Liu, H. Wang, A. Lavin, and D.S. Phoenix: A generative vision model that trains with high data efficiency and breaks text-based CAPTCHAs. Science 10.1126/science.aag2612 (2017)
9. Eddy llg, Nikollaus Mayer, Tonmoy Saikia, Margret Keuper, Alexey Dosovitskiy, and Thomas Brox: FlowNet 2.0: Evolution of Optical Flow Estimation with Deep Networks. arXiv:1612.01925 [cs.CV] Dec. 6, 2016; https://github.com/NVIDIA/flownet2-pytorch
10. Edward Craft, Hartmut Schütze, Ernst Niebur, and Rüdiger von der Heydt: A Neural Model of Figure-Ground Organization. J Neurophysiol 97: 4310-4326, 2007, doi:10.1152/jn.00203.2007
11. Fangtu T. Qiu, Tadashi Sugihara, and Rüdiger von der Heydt: Figure-ground mechanisms provide structure for selective attention. Nat Neurosci. 2007 November; 10(11):1492-1499. doi:10.1038/nn1989
12. Girard, P., J. M. Hupe, and J. Bullier: Feedforward and Feedback Connections Between Areas V1 and V2 of the Monkey Have Similar Rapid Conduction Velocities. J. Neurophysiol. 85, 1328-1331
13. Hee-kyoung Ko, and Rüdiger von der Heydt: Figure-ground organization in the visual cortex: does meaning matter? Articles in Press. J Neurophysiology (Oct 4, 2017) doi:10.1152/jn.00131.2017
14. Janneke F.M. Jehee, Victor A.F. Lamme, and Pieter R. Roelfsema: Boundary assignment in a recurrent network architecture. Vision Research, 47(9), 1153-1165, doi:10.1016/j.visres.2006.12.018
15. Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick: Mask RCNN. arXiv:1703.06870v3 [cs.CV] 24 Jan 2018
16. Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun: Deep Residual Learning for Image Recognition. arXiv:1512.03385v1 [cs.CV] 10 Dec. 2015
17. Magnus Wrenninge, Jonas Unger: Synscapes: A Photorealistic Synthetic Dataset for Street Scene Parsing. arXiv:1810.08705 [cs.CV] Oct.19, 2018; https://7dlabs.com/Synscapes-overview
18. Massimo Minervini, Andreas Fischbach, Hanno Scharr, and Sotirios A. Tsaftaris: Finely-granted annotated datasets for image-based plant phenotyping. Pattern Recognition Letters, doi:10.1016/j.patrec.2015.10.013; https://www.plant-phenotyping.org/datasets-home
19. Mathieu Aubry, Daniel Maturana, Alexei A. Efros, Bryan C. Russell, Josef Sivic: Seeing 3D chairs: exemplar part-based 2D-3D alignment using a large dataset of CAD models.

FlyingChair dataset: https://lmb.informatik.uni-freiburg.de/resources/datasets/GenerateChairs.en.html
20. Olaf Ronneberger, Philipp Fischer, and Thomas Brox: U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv:1505.0457v1 [cs.CV] 18 May 2015
21. Patrik Sundberg, Thomas Brox, Michael Maire, Pablo Arbelaez, and Jitendra Malik: Occlusion Boundary Detection and Figure/Ground Assignment from Optic Flow. *CVPR 2011,* 2011, pp. 2233-2240, doi: 10.1109/CVPR.2011.5995364.
22. Pedro O. Pinheiro, Tsung-Yi Lin, Ronan Collobert, and Piotr Dollár: Learning to Refine Object Segments. arXiv:1603.08695v2 [cs.CV] 26 Jul 2016
23. Pinglei Bao and Doris Y. Tsao: Representation of multiple objects in macaque category-selective areas. Nature Communications | (2018)9.1774, doi:10.1038/s41467-018-04126-7
24. Pinglei Bao, Liang She, Mason McGill and Doris Y. Tsao: A map of object space in primate inferotemporal cortex. Nature Article, published 17 June 2020 doi:10.1038/s41586-020-2350-5
25. Rüdiger von der Heydt: Figure-ground organization and the emergence of proto-objects in the visual cortex. Front.Phychol. 6:1695. doi:10.3389/fpsyg.2015.01695
26. Rüdiger von der Heydt, and Nan R. Zhang: Figure and ground: how the visual cortex integrates local cues for global organization. J Neurophysiol 120: 3085-3098, 2018; published July 25, 2018; doi:10.1152/jn00125.2018
27. Shude D. Zhu, Li Alex Zhang, and Rudiger von der Heydt: Searching for object pointers in the visual cortex. JNP Journal of Neurophysiology, May 11, 2020, doi:10.1152/jn.00112.2020
28. Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, Jianxiong Xiao: 3D ShapeNets: A Deep Representation for Volumetric Shapes. arXiv:1406.5670v3 [cs.CV] Apr. 15 2015; https://modelnet.cs.princeton.edu/
29. Tianlong Chen, Xuemei Cheng, Thomas Tsao: Supplementary Materials for: Border-ownership, Category-selectivity, and Beyond.