Supplementary Material: Manipulating Attributes of Natural Scenes via Hallucination

LEVENT KARACAN, Hacettepe University and Iskenderun Technical University, Turkey
ZEYNEP AKATA, University of Tübingen, Germany
AYKUT ERDEM, Hacettepe University, Turkey
ERKUT ERDEM, Hacettepe University, Turkey

ACM Reference Format:
Levent Karacan, Zeynep Akata, Aykut Erdem, and Erkut Erdem. 2019. Supplementary Material: Manipulating Attributes of Natural Scenes via Hallucination. 1, 1 (October 2019), 18 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

The purpose of this document is to provide extra material to complement the paper. It presents sample annotations for the semantic layouts and the transient attributes that we collected for our dataset (Section 1). It shows additional visual comparisons between the proposed scene generation network (SGN) and Pix2pixHD [Wang et al. 2018], and gives the detailed description of the user study we performed to compare the two methods (Section 2). It explains how our SGN model can be utilized as an image editing tool in which one can play with the layout of the input scene and make changes in its transient attributes (Section 3). Lastly, it demonstrates additional attribute manipulation results on real images (Section 4).

Moreover, we have an accompanying project website[1] which includes

- All 17,772 images from our ALS18K dataset which also contains layout annotations with 150 semantic categories and 40 transient attributes
- An interactive demo of our SGN model
- Additional results on attribute manipulation
- Additional results on season transfer to paintings

1 DATASET COLLECTION

In this section, we discuss our data collection efforts for our ALS18K dataset in more detail. Fig. 1 presents example semantic layout predictions for some images from the Transient Attribute dataset [Laffont et al. 2014], obtained with the method in [Zhao et al. 2017]. In a similar fashion, Fig. 2 illustrates transient attributes estimated by the network in [Baltenberger et al. 2016] for some images from the ADE20K dataset [Zhou et al. 2017]. Finally, Fig. 3 shows the distribution of object classes in our proposed ALS18K dataset, sorted by their number of occurrences.

[1] https://web.cs.hacettepe.edu.tr/~karacan/projects/attribute_hallucination/

Authors' addresses: Levent Karacan, Hacettepe University and Iskenderun Technical University, Ankara, Turkey, karacan@cs.hacettepe.edu.tr; Zeynep Akata, University of Tübingen, Tubingen, Germany, zeynep.akata@uni-tuebingen.de; Aykut Erdem, Hacettepe University, Ankara, Turkey, aykut@cs.hacettepe.edu.tr; Erkut Erdem, Hacettepe University, Ankara, Turkey, erkut@cs.hacettepe.edu.tr.

XXXX-XXXX/2019/10-ART $15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

2 ADDITIONAL COMPARISONS BETWEEN SGN AND PIX2PIXHD

In Fig. 4-5, we present additional comparisons of our SGN model to Pix2pixHD [Wang et al. 2018] method. Note that Pix2pixHD generates scenes conditioned only on the layout but not the transient attributes. Besides enabling control over transient attributes, our proposed SGN model also produces higher quality images than Pix2pixHD.

As an essential part of our evaluation, we had also conducted a user study on Figure Eight, asking subjects to select among the results of our SGN model and the Pix2pixHD method which they think is more photorealistic. We did not set a time limit to the workers to make their decisions. A screenshot of our user interface is shown in Fig. 6. Moreover, Fig. 7 outlines the demographics of our participants. The majority of them were between the ages of 25 and 34 years old, the youngest being 18 and the oldest being 65. The gender ratio was skewed towards males (67% males and 33% females), and most of them have no technical expertise (28% had no specific interest, 61% were hobbyist, 11% were working on image processing/computer graphics.

3 A STANDALONE GUI PROTOTYPE

We designed a GUI prototype that provides users with the ability to use our approach as an interactive photo editing tool. Fig. 8 presents a screenshot of our application. The GUI the layouts shows the input image and its semantic layout and it has controls to synthesize different versions of the scene based on desired set of transient attributes. Moreover, it allows users to control the degree of the attributes via the "increase" and "decrease" buttons. We also note that one can generate slightly different versions of the hallucinated scene by playing with the "Random Noise" button. Additionally, one can play with the layout to generate a novel scene from scratch. In `demo-gui.mp4`, we provide a video of the GUI in action.

4 ADDITIONAL RESULTS FOR AUTOMATIC ATTRIBUTE MANIPULATION

In our article, we have only provided attribute manipulation results obtained by using FPST [Li et al. 2018] as it often gives more favorable results than DPST [Luan et al. 2017]. In Fig. 9-12, for the sake completeness, we present here comparison results with previous work where we employ DPST for transferring the attributes

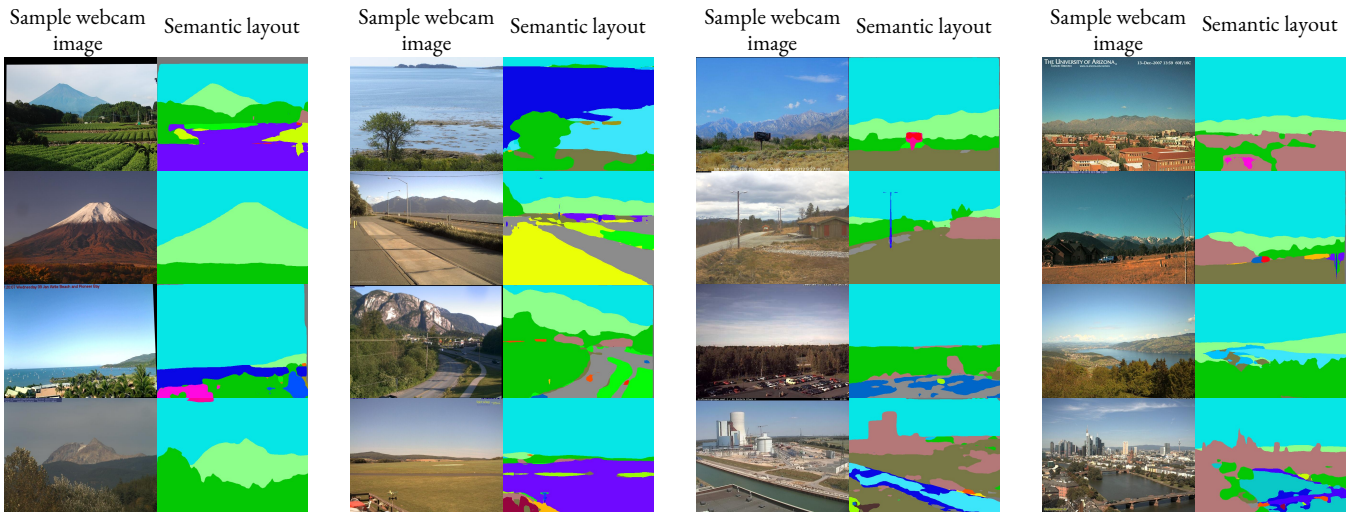

Fig. 1. Sample semantic layout predictions for some images from Transient Attribute dataset Laffont et al. [2014].

of the hallucinated images to the input images within our framework. Moreover, we show additional attribute manipulation results in Fig. 13-15. For each natural scene image given as input, we randomly select a set of transient attributes and generate new versions of the same scene by playing with these attributes.

For evaluating the overall performance of our framework, we had also conducted a second user study on Figure Eight. We showed the subjects an input image and a pair of manipulation results along with a target attribute and let them select one of the manipulated images which they consider visually more appealing regarding the specified target attribute. The manipulated images were the results of our framework obtained with DPST or FPST, or those obtained by the approach by Laffont et al. [2014]. The users have unlimited time to make a selection. A screenshot of our user interface is shown in Fig. 16. Fig. 17 summarizes the demographic distribution of the participants. The majority of them were between the ages of 25 and 34 years old, the youngest being 18 and the oldest being 59. The gender ratio was skewed towards males (34% males and 66% females), and most of them have no technical expertise (34% had no specific interest, 61% were hobbyist, 5% were working on image processing/computer graphics.

REFERENCES

Ryan Baltenberger, Menghua Zhai, Connor Greenwell, Scott Workman, and Nathan Jacobs. 2016. A fast method for estimating transient scene attributes. In *Winter Conference on Application of Computer Vision (WACV)*.

Pierre-Yves Laffont, Zhile Ren, Xiaofeng Tao, Chao Qian, and James Hays. 2014. Transient attributes for high-level understanding and editing of outdoor scenes. *ACM Transactions on Graphics (TOG)* 33, 4 (2014).

Yijun Li, Ming-Yu Liu, Xueting Li, Ming-Hsuan Yang, and Jan Kautz. 2018. A closed-form solution to photorealistic image stylization. In *European Conference on Computer Vision (ECCV)*. 453–468.

Fujun Luan, Sylvain Paris, Eli Shechtman, and Kavita Bala. 2017. Deep photo style transfer. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. 2018. High-resolution image synthesis and semantic manipulation with conditional GANs. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 1–13.

Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. 2017. Pyramid scene parsing network. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Bolei Zhou, Hang Zhao, Xavier Puig, Sanja Fidler, Adela Barriuso, and Antonio Torralba. 2017. Scene parsing through ADE20K dataset. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

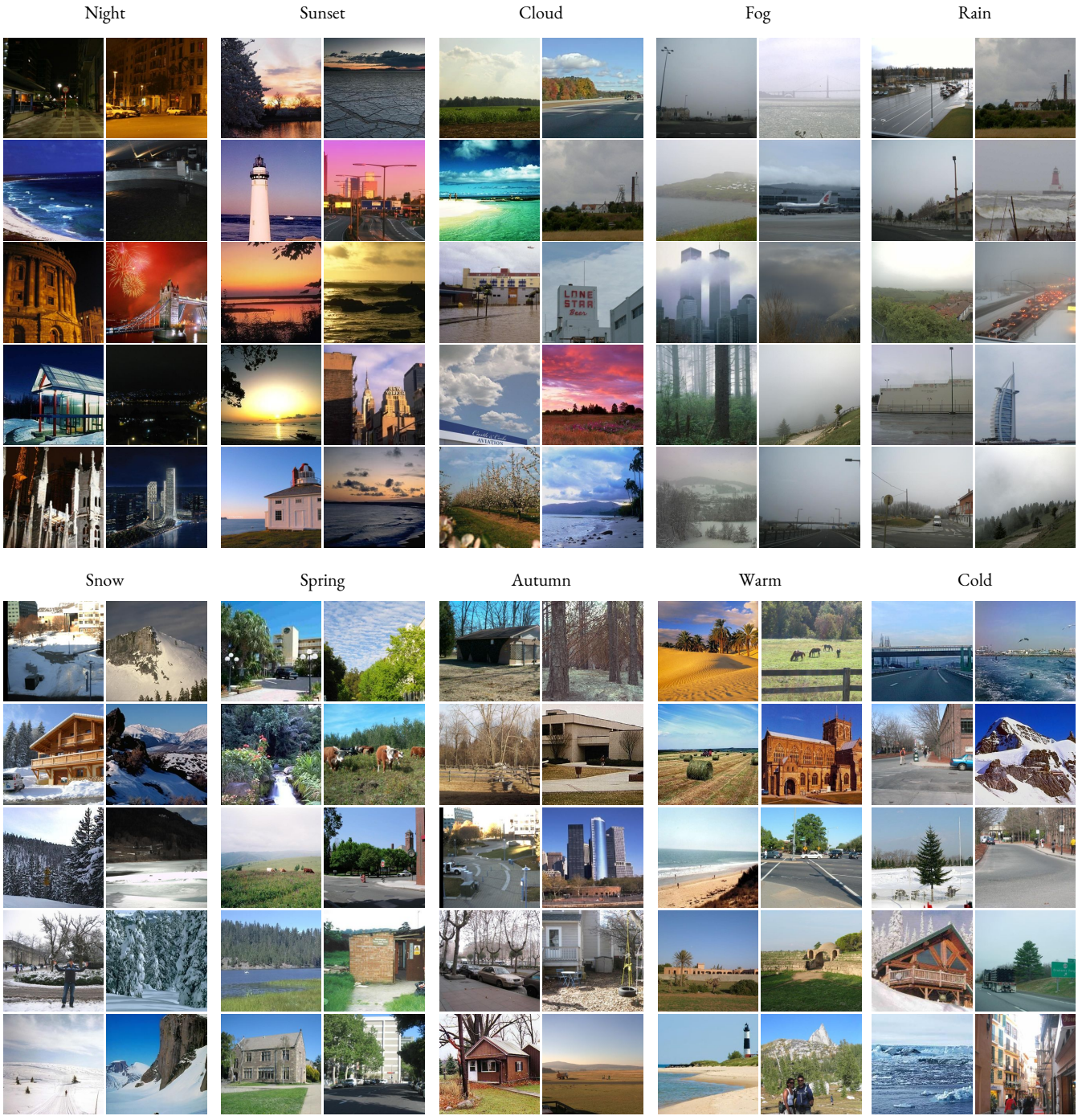

Fig. 2. Sample transient attribute predictions for some images from ADE20K dataset [Zhou et al. 2017].

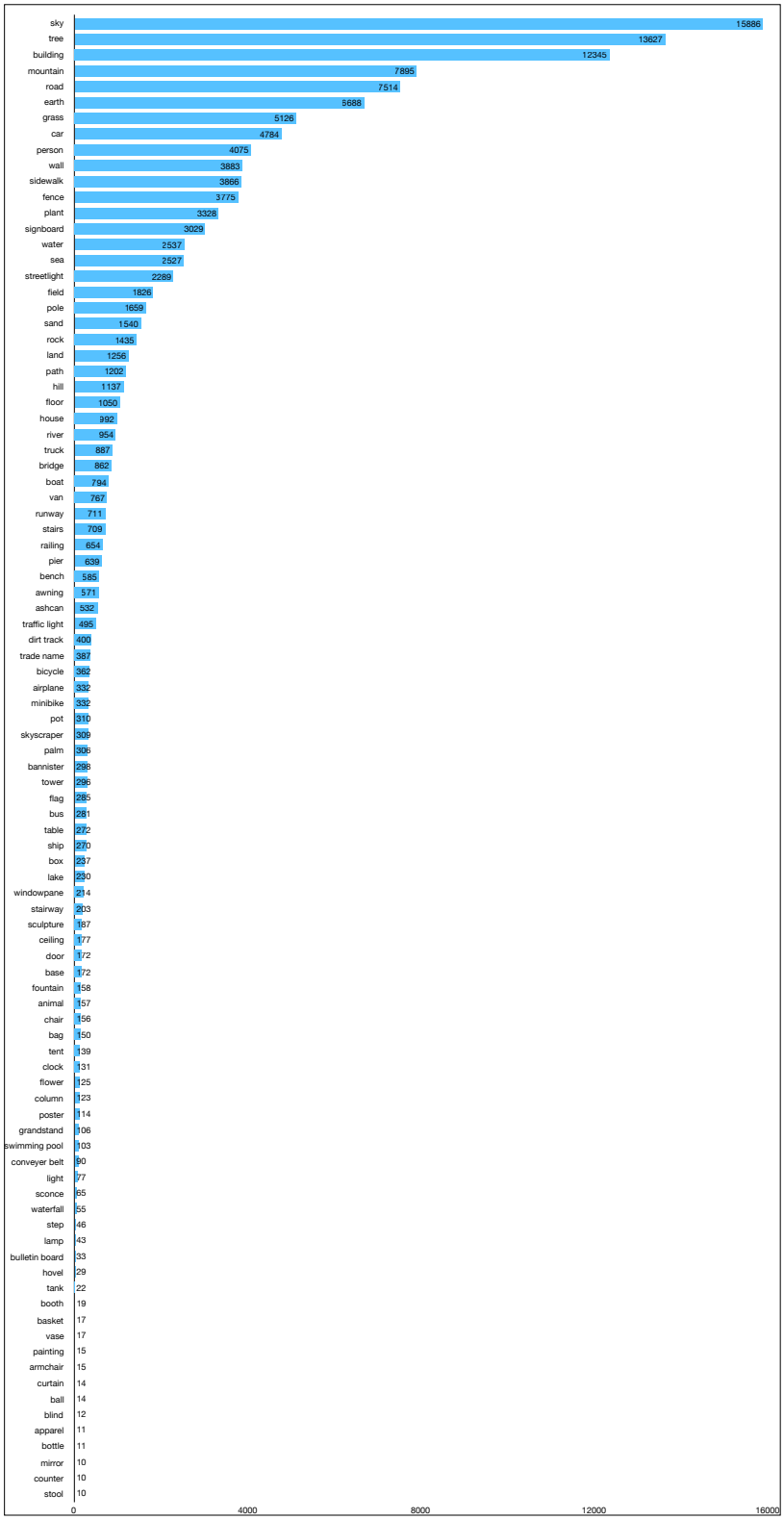

Fig. 3. Distribution of object classes in our proposed ALS18K dataset (sorted by their number of occurrences).

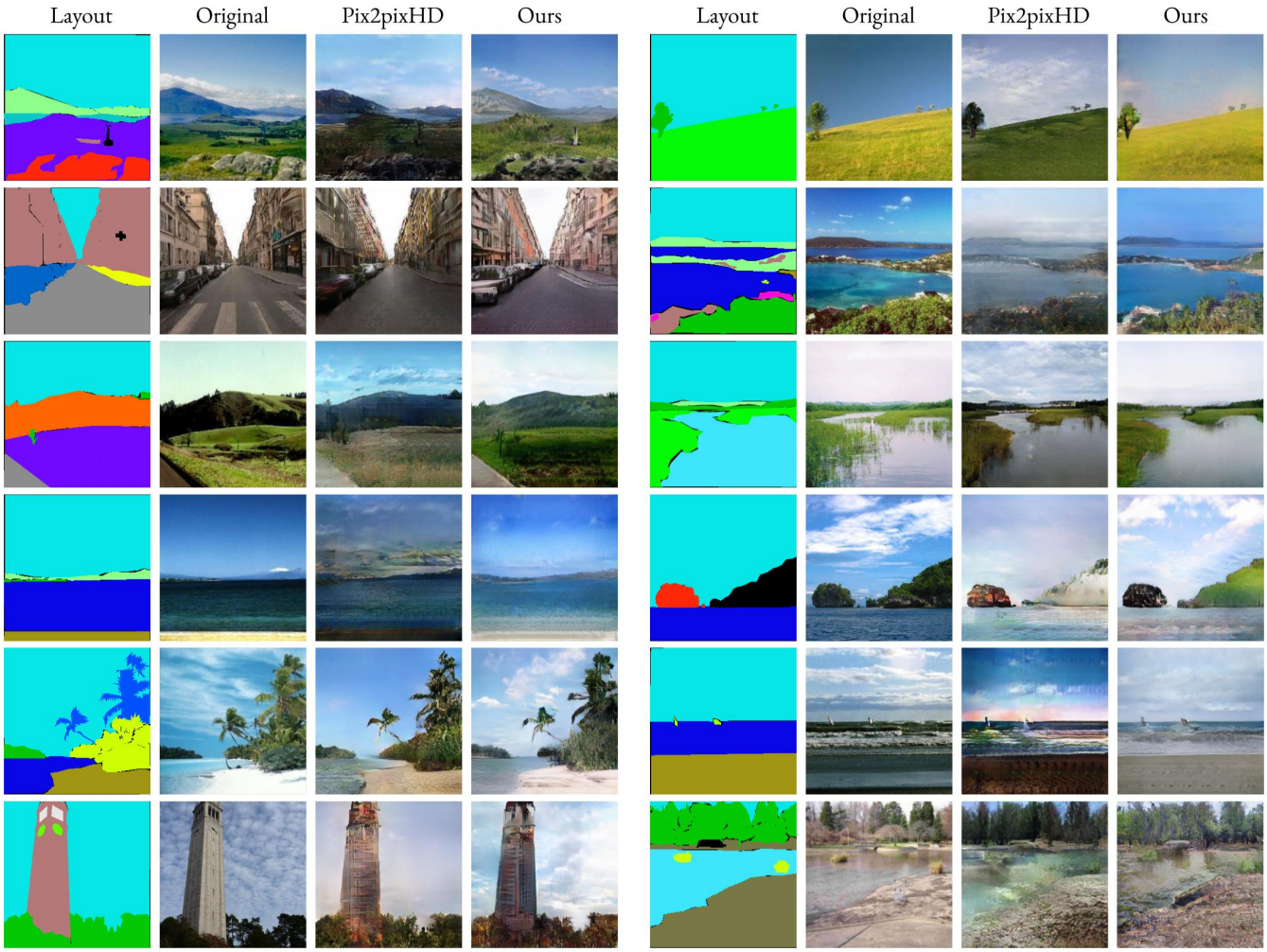

Fig. 4. Comparison of the proposed SGN against Pix2pixHD. Given the input semantic layout, we show the reference image this semantic layout belongs to and the synthetic images generated by Pix2pixHD and our SGN model.

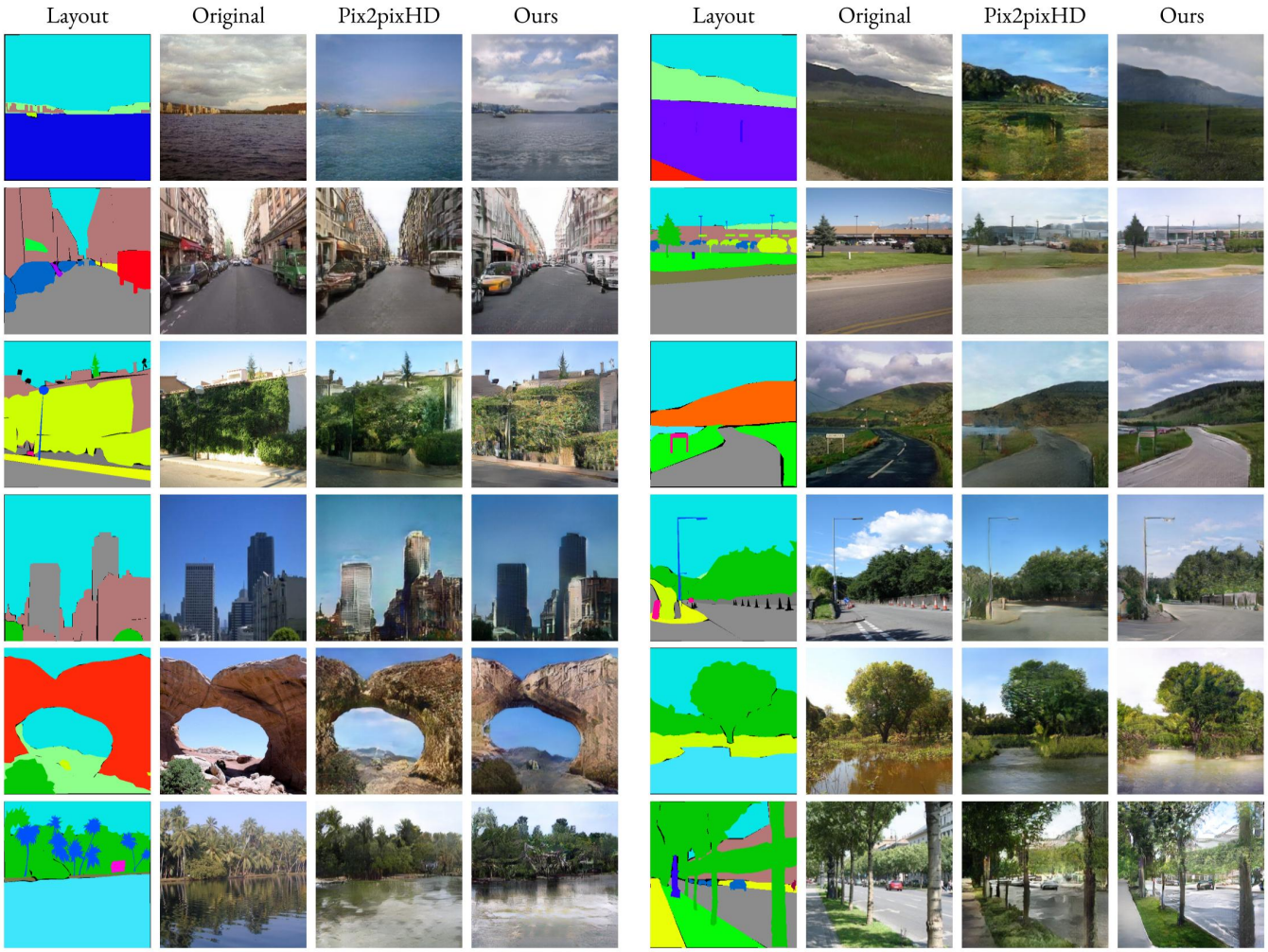

Fig. 5. Comparison of the proposed SGN against Pix2pixHD. Given the input semantic layout, we show the reference image this semantic layout belongs to and the synthetic images generated by Pix2pixHD and our SGN model.

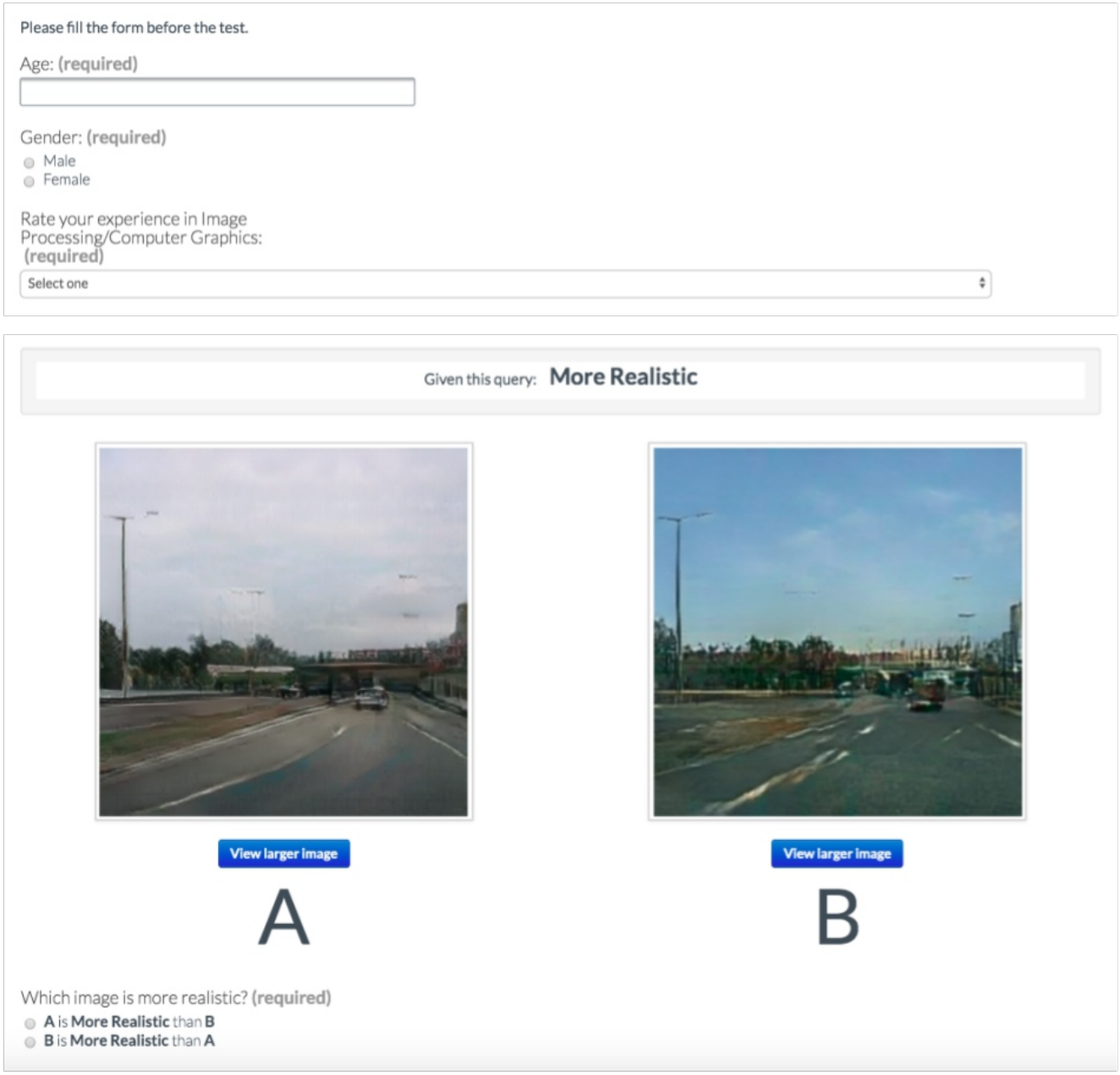

Fig. 6. A screenshot of a sample question in our user study on scene synthesis.

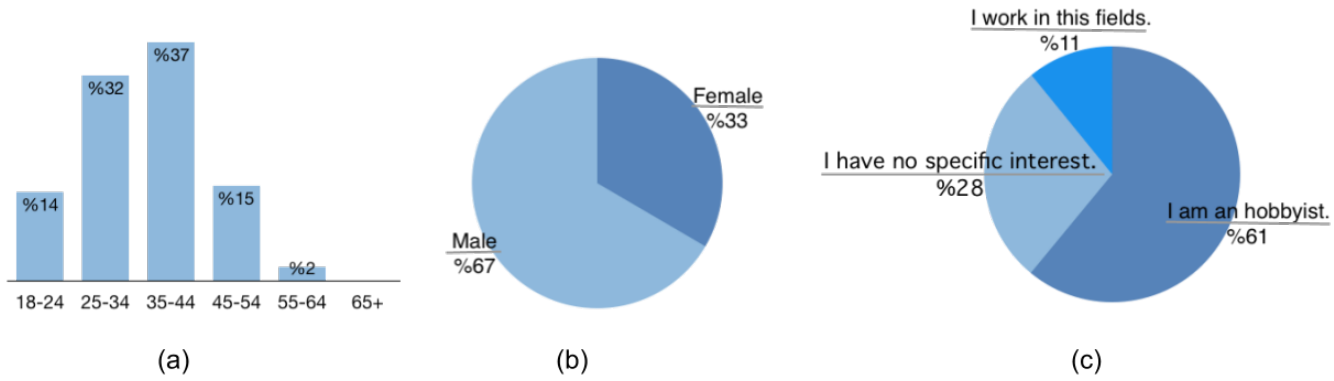

Fig. 7. (a) Age, (b) gender and (c) profession distribution of the participants of our user study on scene synthesis.

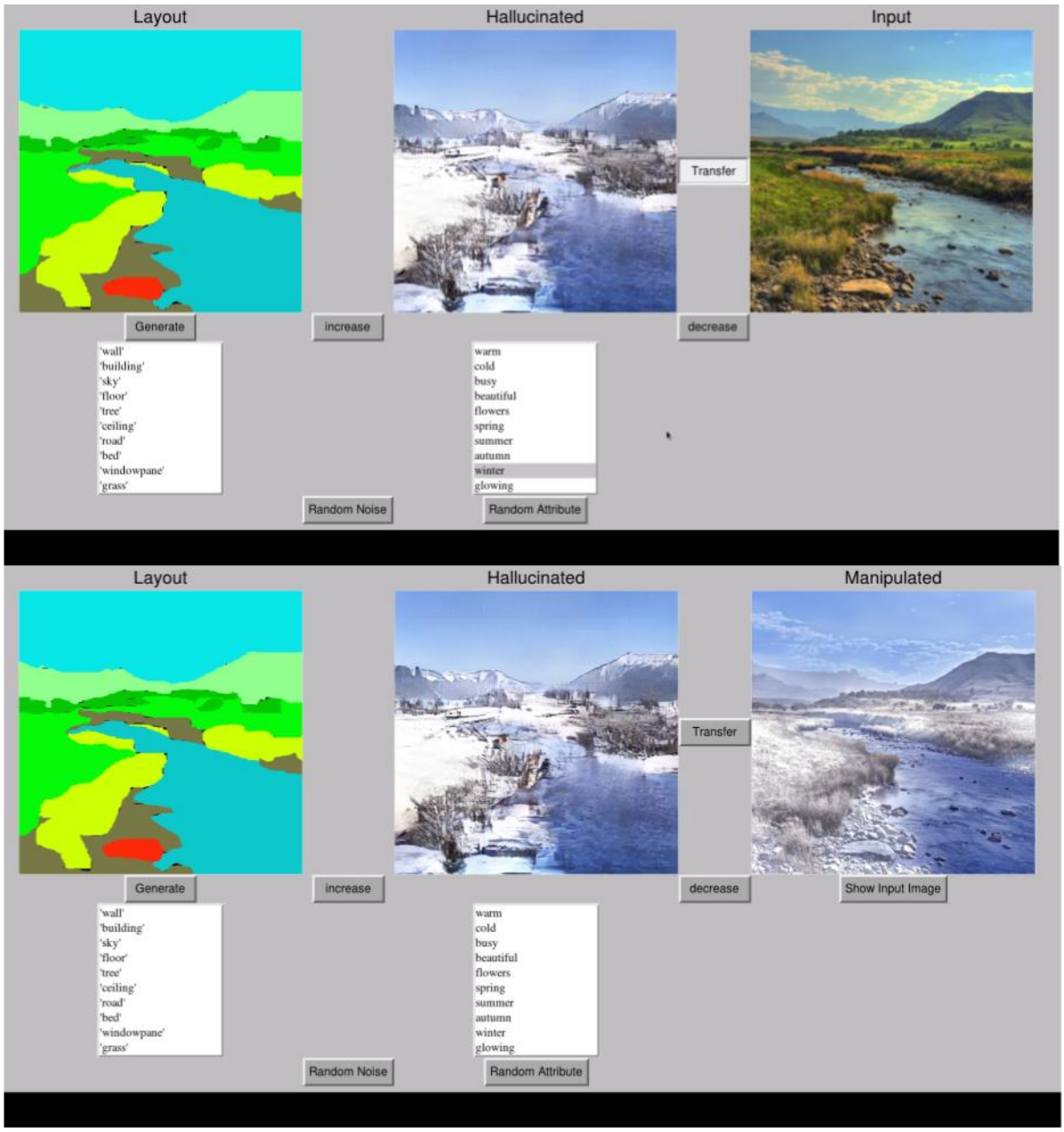

Fig. 8. A screenshot of the GUI prototype running our approach as a backend. See `demo-gui.mp4` for a video of the GUI in action.

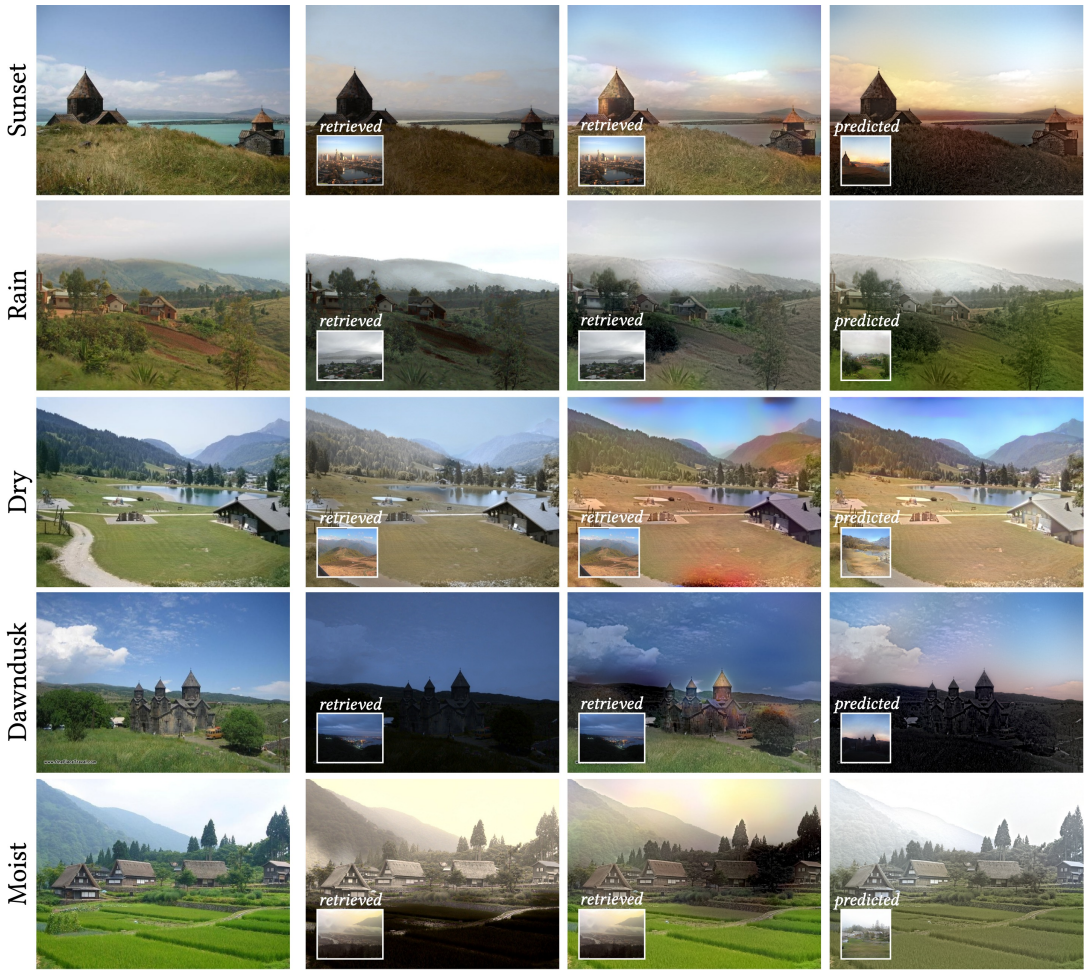

Fig. 9. Comparison with base method of [Laffont et al. 2014]. For given input image, (a) and (b) results of [Laffont et al. 2014] using the exemplar-based style transfer method they proposed and DPST method [Luan et al. 2017] respectively between retrieved images and input images, (c) results of our method using DPST [Luan et al. 2017] between generated image by proposed SGN model and input image.

Fig. 10. Our method with DPST also produces photorealistic manipulation results for different degrees of transient attributes.

Fig. 11. Season transfer to paintings with DPST. Source images: Wheat Field with Cypresse by Vincent van Gogh (1889), In the Auvergne by Jean-Francois Millet (1869) and Lourmarin by Paul-Camille Guigou (1868), respectively.

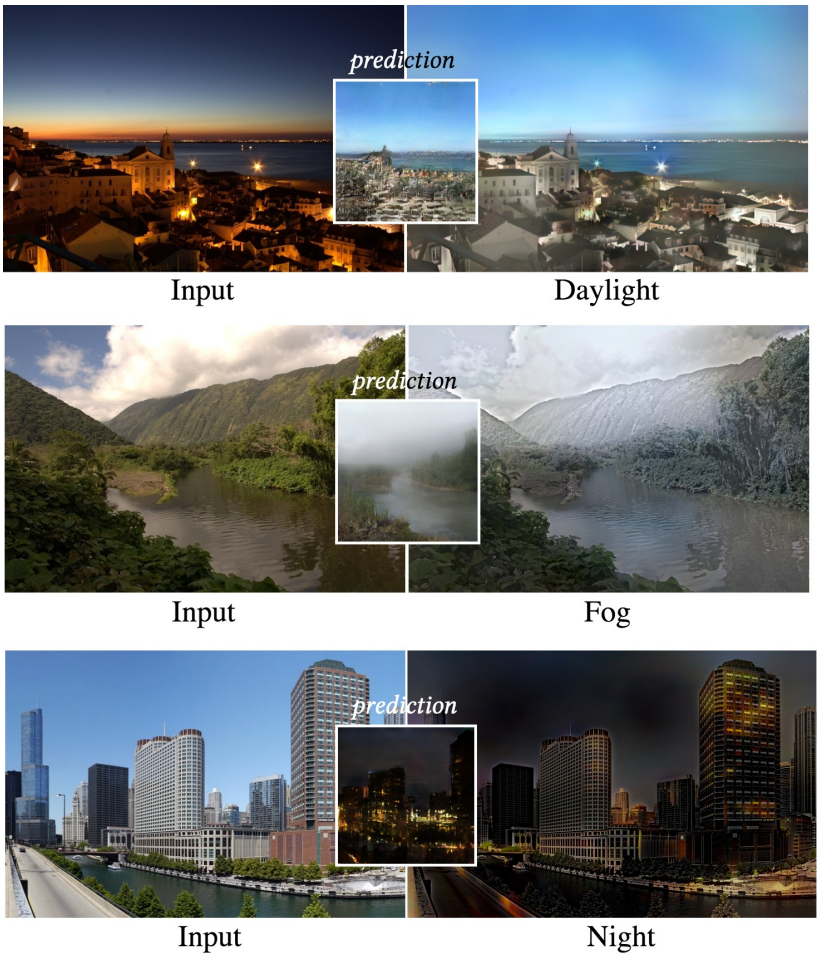

Fig. 12. Example failure cases for our attribute manipulation framework, which are due to the visual quality of synthesized reference style image (top row) and failing of the photo style transfer method (bottom two rows).

Fig. 13. Attribute manipulation results obtained with our framework by using DPST [Luan et al. 2017] and FPST as the style transfer method. Given a natural scene image as input, we select a set of transient attributes and consequently obtain different images of the same scene reflecting the characteristics of these attributes.

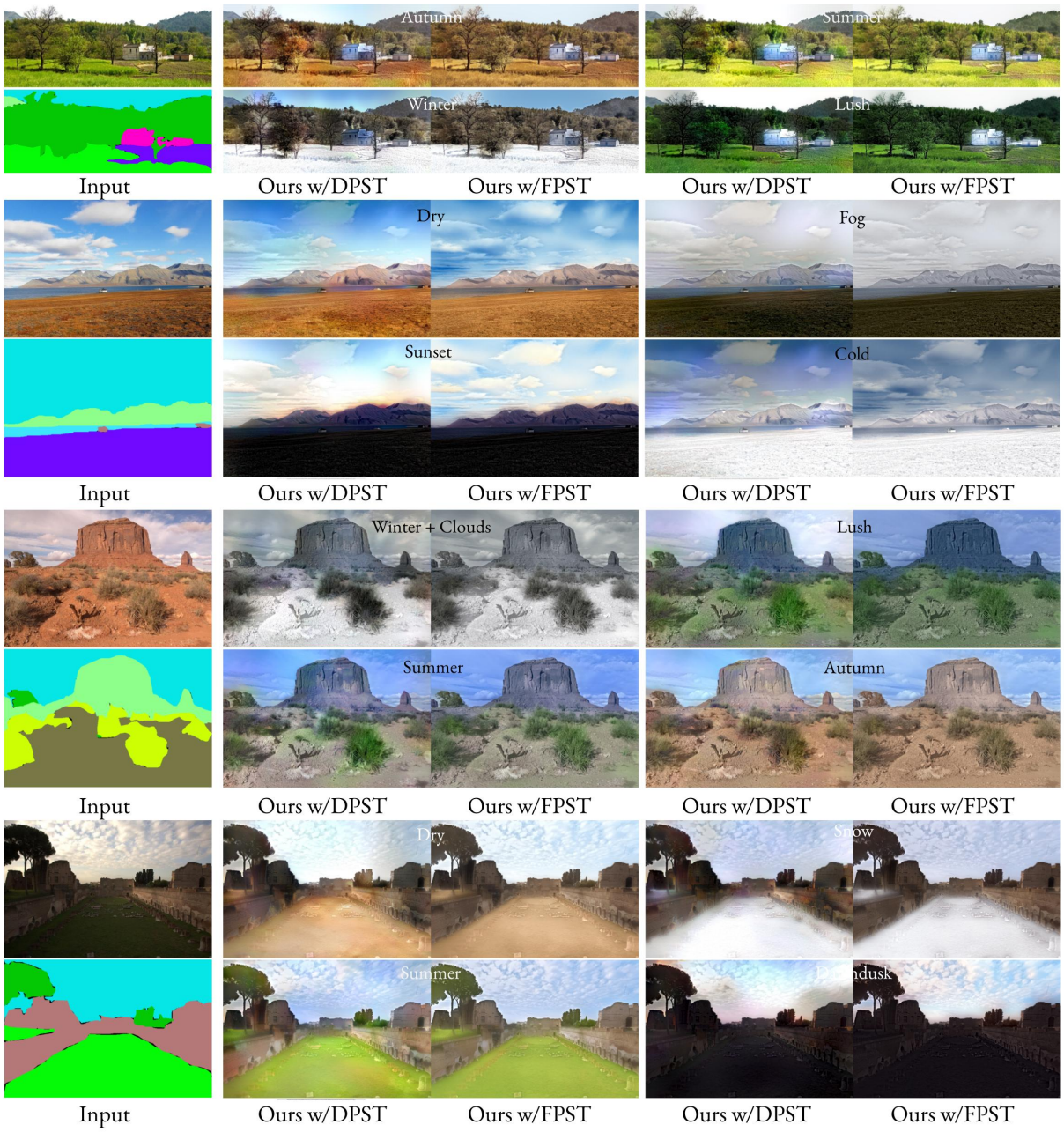

Fig. 14. Attribute manipulation results automatically generated by our framework by using DPST [Luan et al. 2017] and FPST [Li et al. 2018] as the transferring mechanism. Given a natural scene image as input, we select a set of transient attributes and consequently obtain different images of the same scene reflecting the characteristics of these attributes.

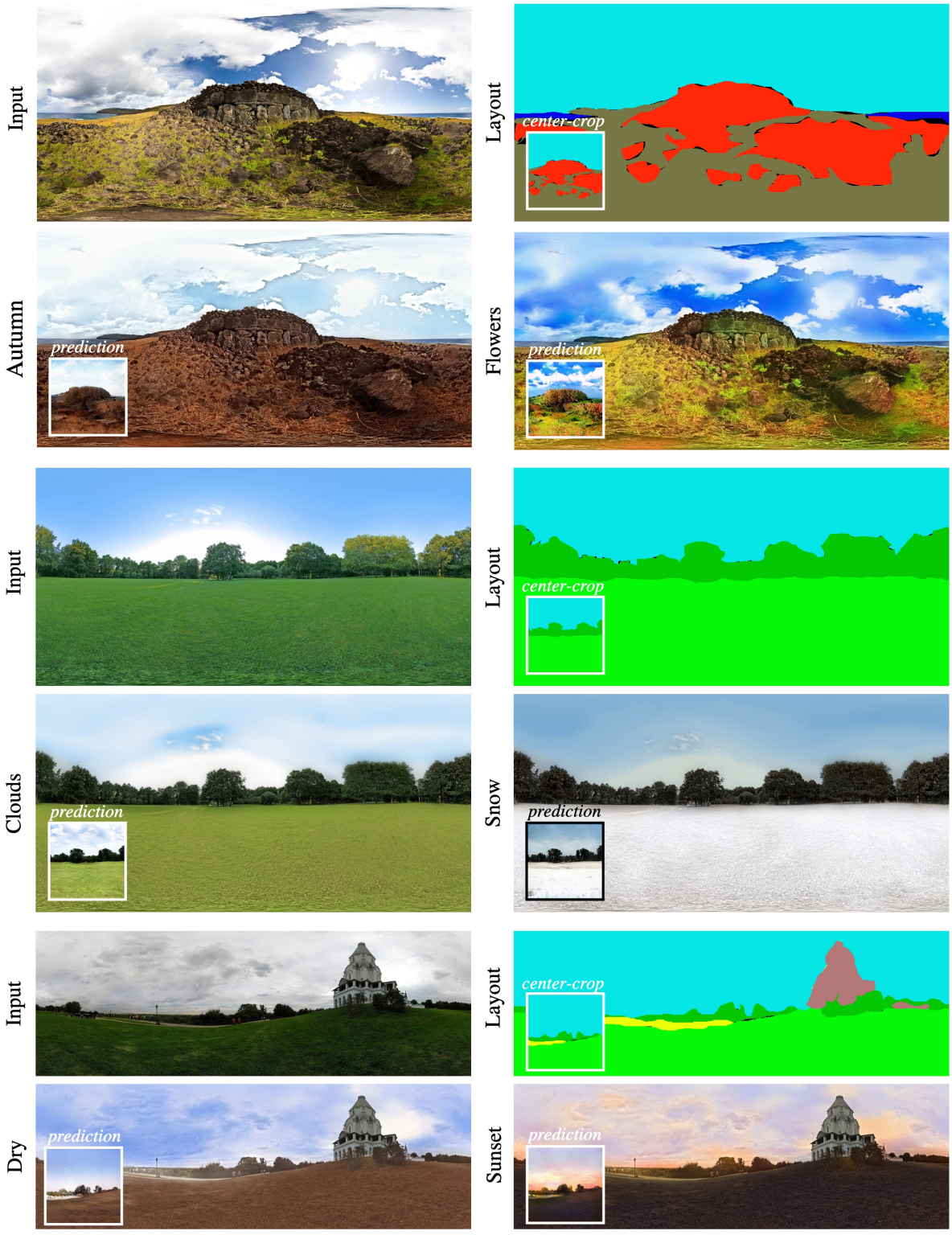

Fig. 15. Attribute manipulation results on panorama images.

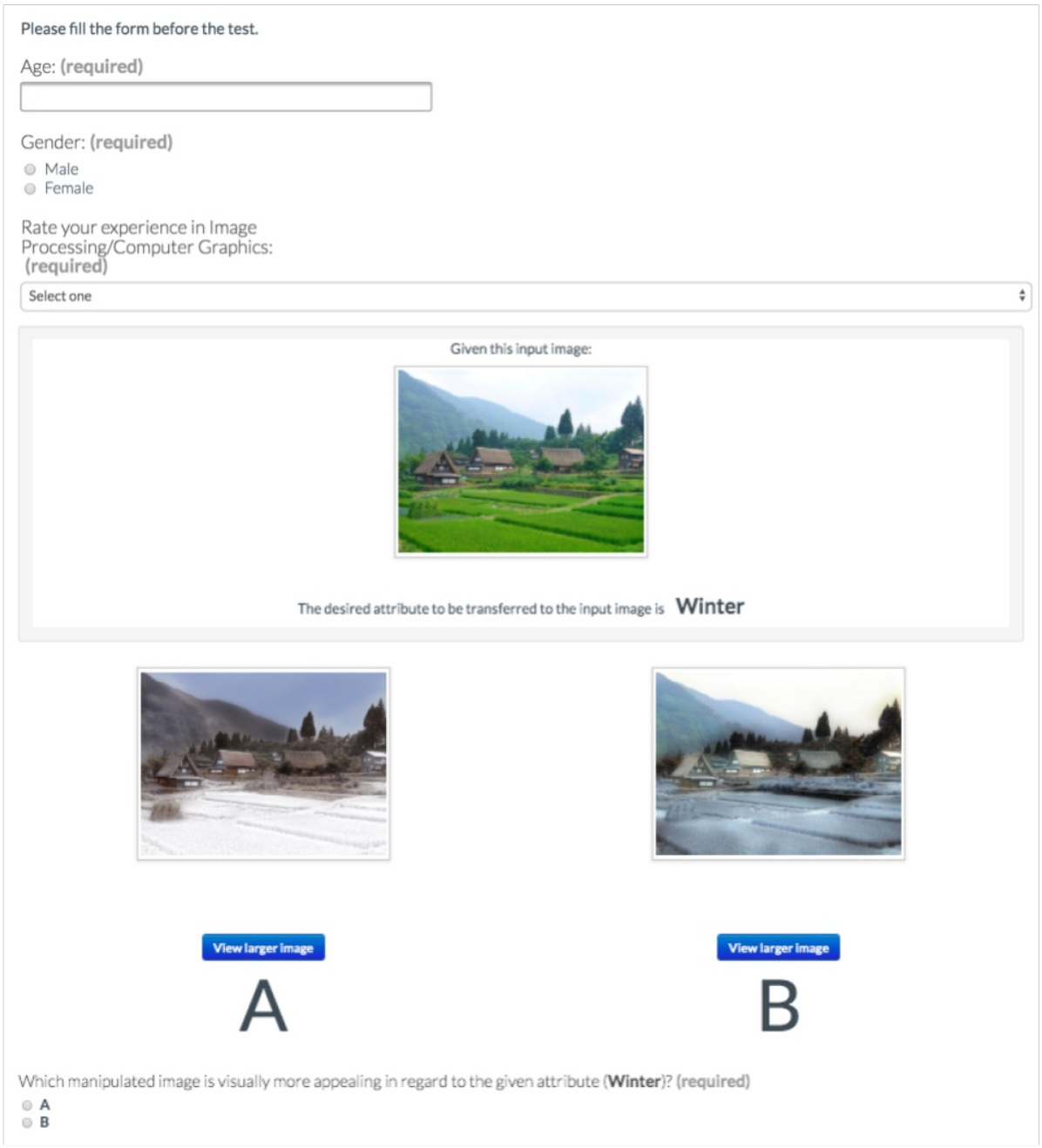

Fig. 16. A screenshot of a sample question in our user study on attribute transfer.

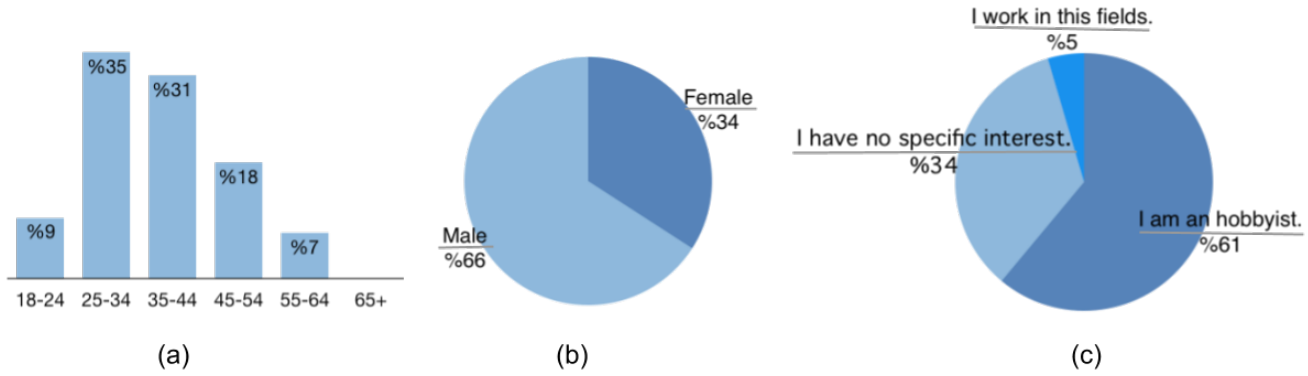

Fig. 17. (a) Age, (b) gender and (c) profession distribution of the participants of our user study on scene synthesis.