

Multimodal Transfer: A Hierarchical Deep Convolutional Neural Network for Fast Artistic Style Transfer, Supplementary Material

Xin Wang^{1,2}, Geoffrey Oxholm², Da Zhang¹, Yuan-Fang Wang¹

¹University of California, Santa Barbara, CA

²Adobe Research, San Francisco, CA

{xwang, dazhang, yfwang}@cs.ucsb.edu, oxholm@adobe.com

Luminance Branch

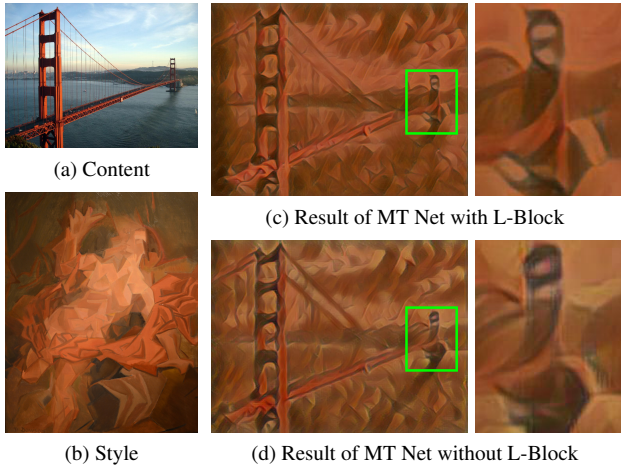


Figure A: Comparison of results with/without L-Block

In Fig. A we demonstrate the effectiveness of the luminance branch *L-Block*. In (c) we show the result with the block, and in (d) when it is omitted. In addition to overall color accuracy (0.904 vs 0.797), the zoom-in areas (right) show how the result with the L-Block is smoother in transition and lacks the jitter artifact that appears when this block is omitted.

Multimodal Transfer Examples

Two additional figures are provided to demonstrate the superior performance of multimodal transfer and compare it to the state-of-the-art network by Johnson *et al.* (Johnson Net). In Fig. B, we show multimodal transfer results for many different styles on a large variety of contents. In Fig. C, we evaluate multimodal transfer on larger images (1800×1352) and compare the results with those of Johnson Net, illustrating the advantages of multimodal transfer in simulating both high-level texture and fine detailed brushwork of the original style guides.

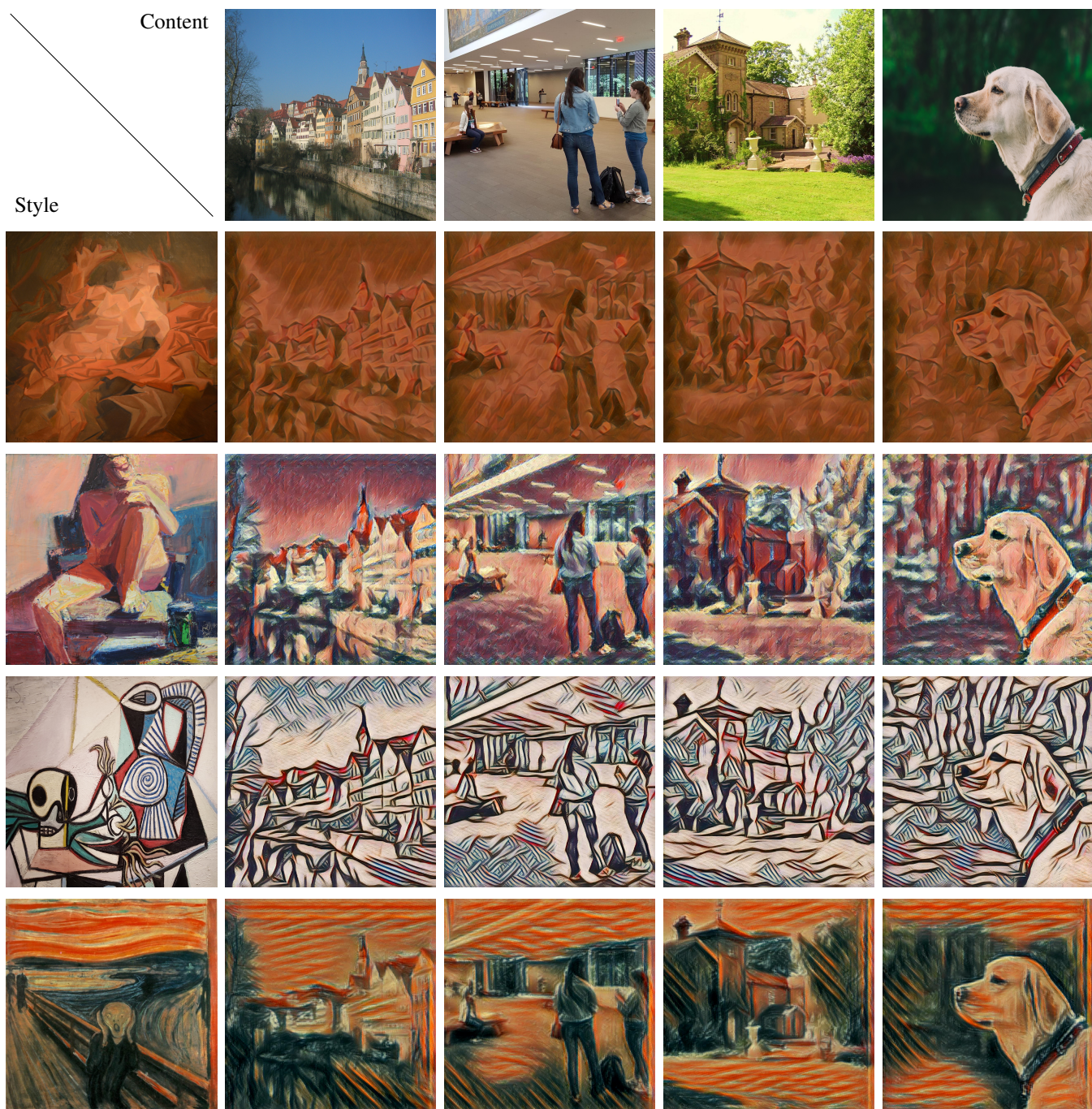
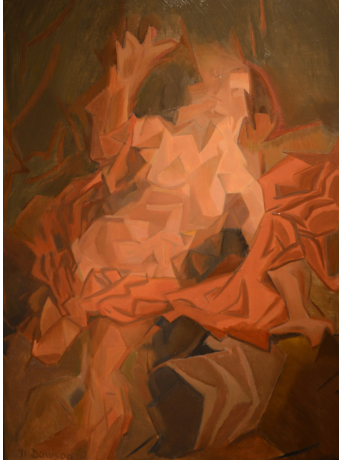


Figure B: Multimodal transfer results for many different artistic styles. The results shown here are organized as a 5×5 grid: The top row shows four content images and the left column shows four different style guides. A particular entry in the grid depicts the result of applying the style guide in the corresponding row to the content image in the corresponding column. All images shown are of size 1024. These examples illustrate that our algorithm works on a large variety of contents—indoors, outdoors, natural landscape, man-made structures, plants and animals—and styles. The style guides shown in the left column, from top to bottom, are *Venus* by Manierre Dawson, *Woman Reflecting* by Terry St. John, *Still Life with Skull, Leeks and Pitcher* by Pablo Picasso, and *The Scream* by Edvard Munch.



(a) Style: *Venus* by Manierre Dawson



(b) Style: *Still Life with Skull, Leeks and Pitcher* by Pablo Picasso



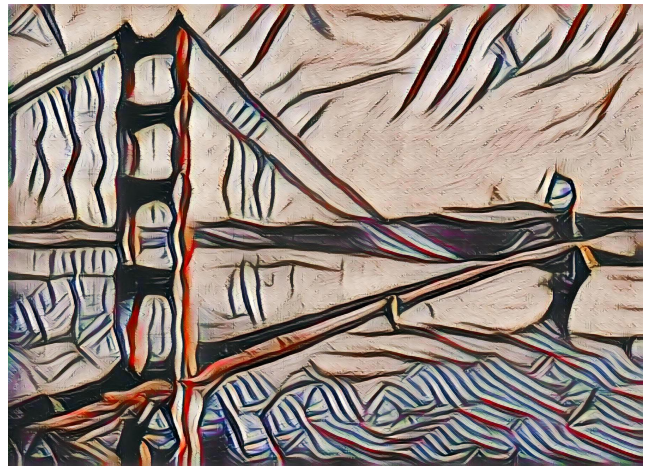
(c) Johnson *et al.*



(d) Johnson *et al.*



(e) Multimodal transfer (Ours)



(f) Multimodal transfer (Ours)

Figure C: Comparison examples on larger images (1800×1352). Here we directly compare our multimodal transfer with the state-of-the-art network by Johnson *et al* (Johnson Net). The first row are two artistic style images used for the experiments; the second row are the corresponding results generated by Johnson Net; and third row are multimodal transfer results generated by our MT Net. All results are 1800×1352 pixels. As can be seen in these two examples, although Johnson *et al*'s results are visually appealing, our results preserve the correct texture scales and are more similar to the original artistic styles both in coarse texture and fine brushwork.