Supplementary Material of Illustration2Vec: A Semantic Vector Representation of Illustrations

Masaki Saito*
Tohoku University

Yusuke Matsui*
The University of Tokyo

1 Additional experiments

Due to limitations of space in our main paper, we show the additional examples of semantic morphing in Fig.1 and 2. The results are almost the same as Fig.5 in the main paper. When we pass the two monochrome line arts to our system, it retrieves monochrome illustrations that smoothly change the attributes from one to the other as intermediate illustrations (the 5th column from the right in Fig.1). If we pass both male and female illustrations as input images, the system tends to retrieve illustrations that combine the both features, i.e., androgynous illustrations. As with the experiment of semantic morphing in the main paper, we carefully selected the two illustrations as input images such that our system returns 4 intermediate illustrations.

2 Training of Convolutional Neural Networks

This section describes the detail of the training of our Convolutional Neural Network (CNN).

2.1 Data augmentation

As with Simonyan *et. al.* [2015], we employ the data augmentation and the image preprocessing algorithms. Before sending an image, we resize it to 256×256 pixels, randomly sample a 224×224 pixels, and randomly flip the images horizontally with 50% probability. Then, we subtract the mean of all the training images from raw pixel values, and send it into our network.

2.2 Training settings

We use a Stochastic Gradient Descent (SGD) method to optimize the proposed model. We set the momentum to 0.9 and the weight decay to 5.0×10^{-4}. For regularization we respectively insert two dropout layers before two 1024-channel convolutional layers. The dropout ratio is set to 0.5. The network is initialized with learning rates of 1.0×10^{-3}, and this value is further reduced by hand whenever the test error stops improving. We split the dataset by assigning 90% of the images to the training set, and 10% to a test set. We use caffe [Jia 2013] on a single machine with GTX Titan Black. The training itself takes approximately 7 days.

3 References of illustrations

In order to illustrate the qualitative results, we show several illustrations in our main paper with links. A source of an image in the main paper and the supplementary material can be seen by clicking the image.

References

JIA, Y., 2013. Caffe: An Open Source Convolutional Architecture for Fast Feature Embedding.

SIMONYAN, K., AND ZISSERMAN, A. 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. In *ICLR*, Computational and Biological Learning Society.

*Authors contributed equally

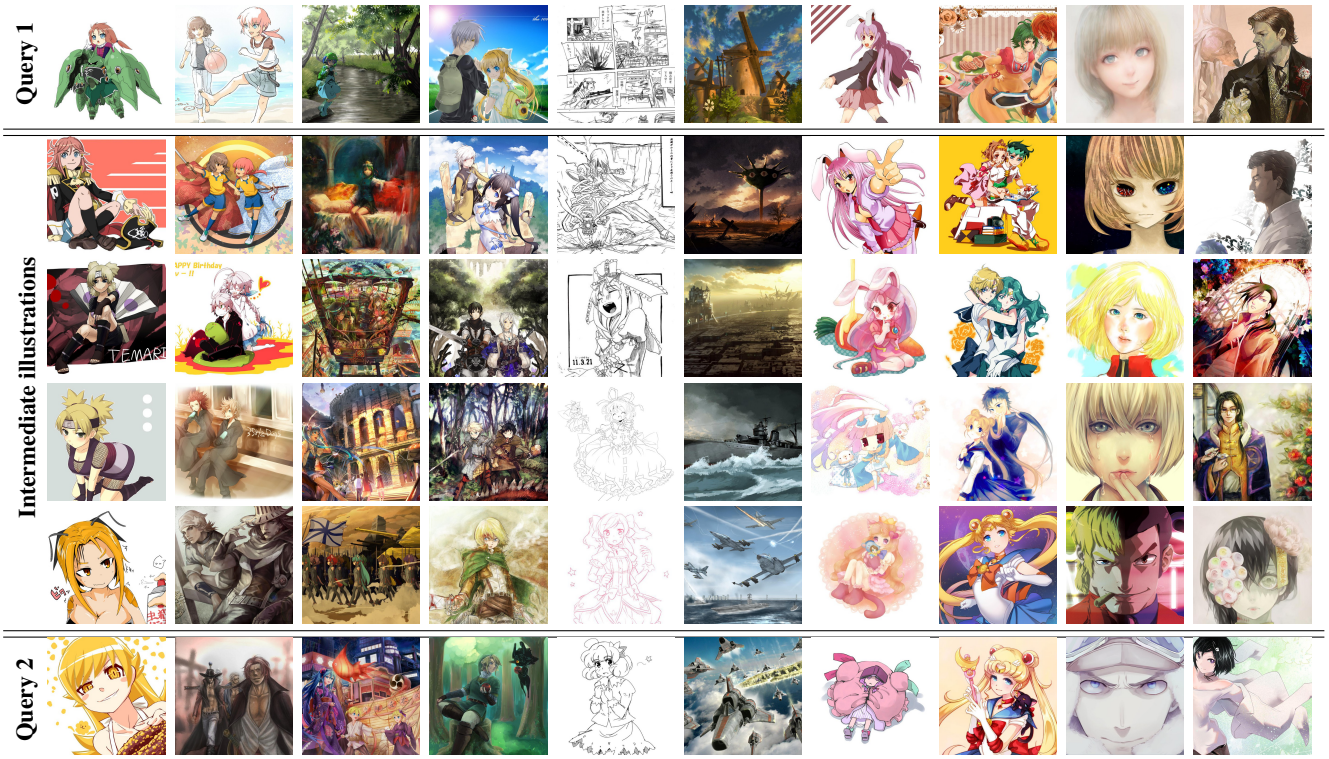

Figure 1: *Additional examples of semantic morphing.*

Figure 2: *Additional examples of semantic morphing.*