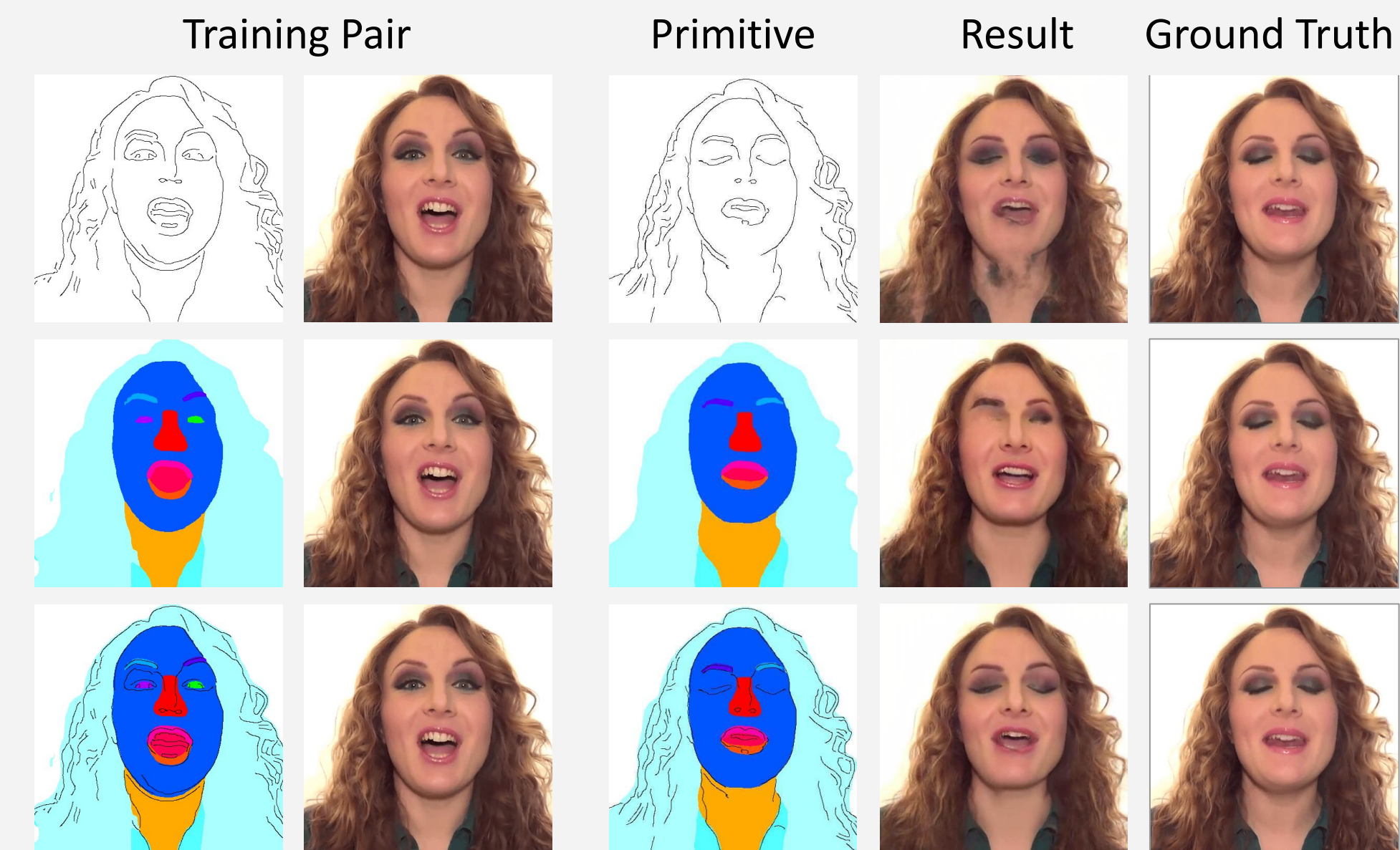


Abstract

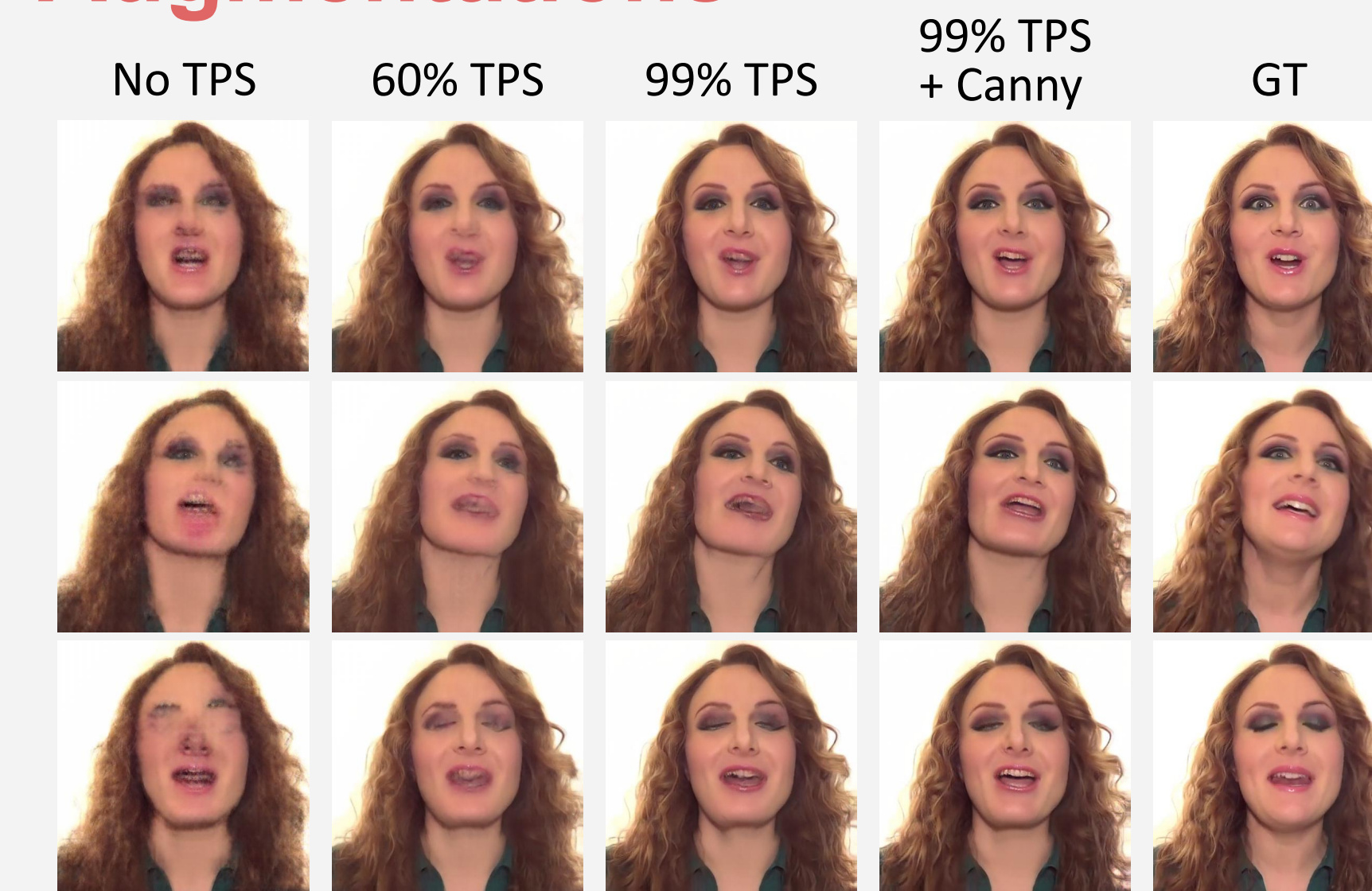
We propose a simple-to-implement yet highly effective method for training deep conditional generative models from a single image pair

Primitive Representation

Using a combination of edges and segmentation maps allows the model to capture the fine details as well as the semantics of the images



Augmentations

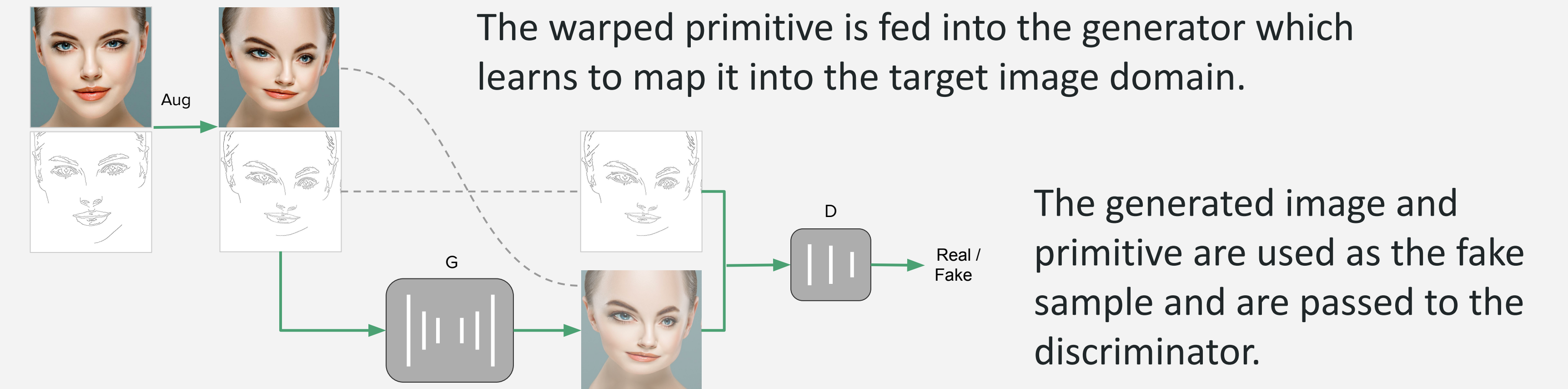


The main augmentation we use is TPS. However, additional augmentations improve sharpness



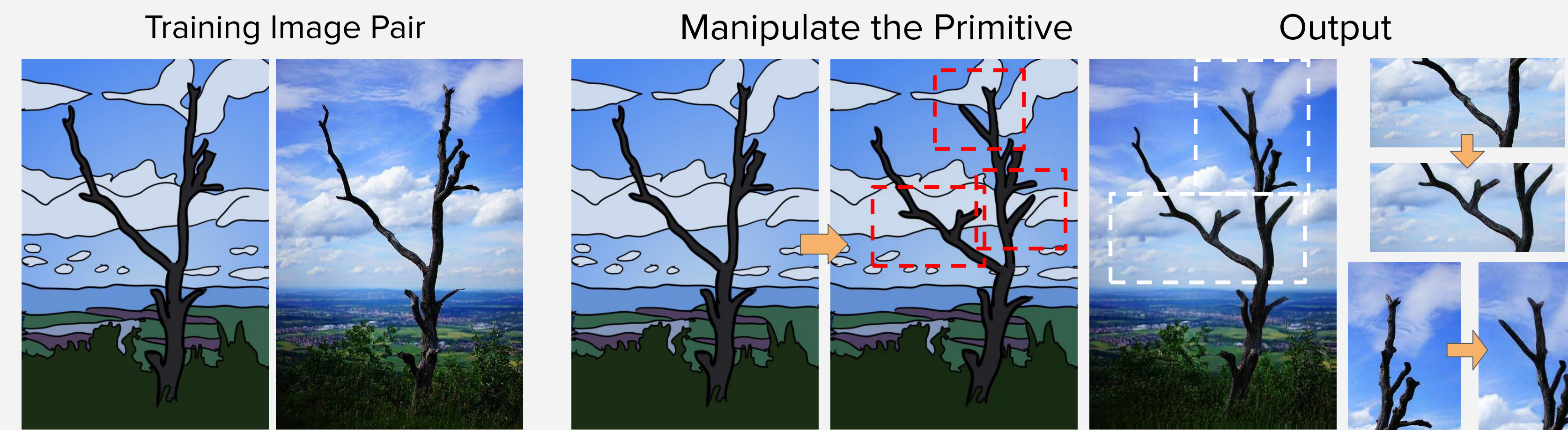
Training Pipeline

- (1) Given the input image, a simpler representation of the image is created, which we dub the "primitive".
- (2) Next, we augment both the target and primitive images.
- (3) We then use a standard conditional image-to-image mapping network to learn a mapping between the primitive representation and the image. This process is repeated at every iteration.

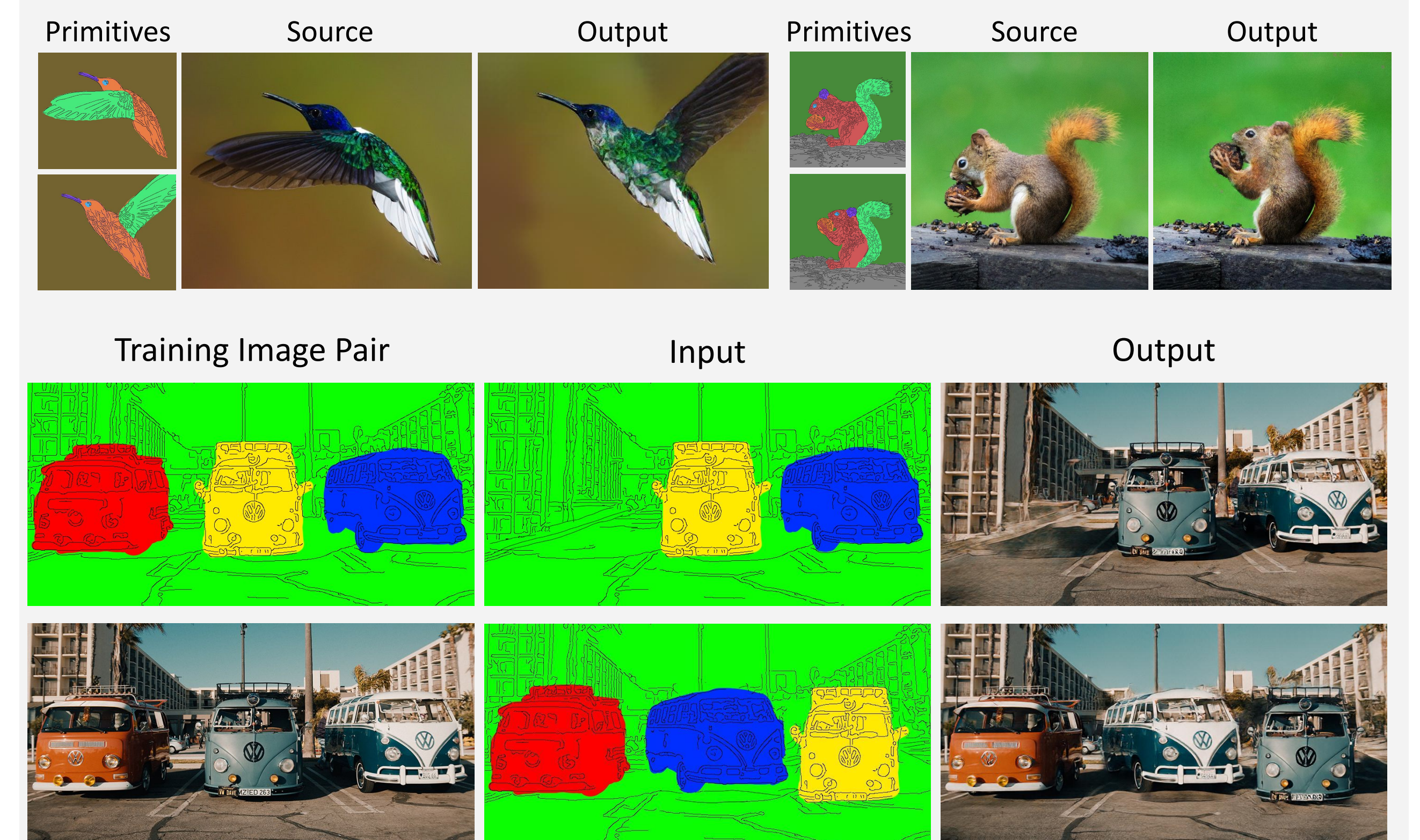


Inference

At inference, the original primitive is manipulated by the user, changes highlighted in red. Then, the manipulated primitive is passed through the network which outputs a corresponding manipulated image in the real image domain. On the right, we can see that the manipulation was performed successfully, while preserving the internal statistics of the source image



Additional Results



Single Image Animation

By training only on a single image-primitive pair, DeepSIM is able to translate a short video clip from one image domain to the other.

Animation to photorealistic:



Photorealistic to Animation:

