This is another advantage and difference respect to the other approaches used by other reaserchers

Additionally, our system also goes beyond state-of-the-art in that it can handle images under

changing backgrounds and illumination conditions. We achieve this by means of

an attention layer that focuses the action of the network only in those regions

of the image that are relevant to convey the novel expression.

This increases the power of the model that it is not only able to create expressions in a continue domain but also to handle images in the wild with complex backgrounds and illumination conditions (thanks to attention mechanism) – the attention mask learns what are the pixels that are important for a specific AU.

How GAN works? There is a generator network that learns how to build realistic fake samples and a discriminator network that learns how to distinguish between real and fake data.

The generator G(Iyr jyg) is trained to realistically transform the facial expression

in image Iyr to the desired yg.

One key ingredient of our system is to make G focus only on those regions

of the image that are responsible of synthesizing the novel expression and keep

the rest elements of the image such as hair, glasses, hats or jewelery untouched.

For this purpose, we have embedded an attention mechanism to the generator.

Our generator outputs two masks, a color mask C and attention mask A.

The mask A indicates to which extend each pixel of the C contributes to the output

image Iyf . In this way, the generator does not need to render static elements,

and can focus exclusively on the pixels defining the facial movements, leading to

sharper and more realistic synthetic images.

Given an input image and the target expression,

the generator regresses and attention mask A and an RGB color transformation C over

the entire image. The attention mask defines a per pixel intensity specifying to which

extend each pixel of the original image will contribute in the final rendered image.

We have to include an attention loss to the final loss to drive the attention masks to be smooth and prevent it from saturating. In fact it has seen that the attention mask can easily saturate and this turn into no effects produced by the generator (input is the same of the output) – the attention mask isn’t able to see relevant part of the images so the generator just copyes the entire image – the solution is to regulize the mask with a l2-weight penaloty.

Fig. 5 displays, for the same experiment, the attention A and color C masks

that produced the final result Iyg . Note how the model has learned to focus its

attention (darker area) onto the corresponding AU in an unsupervised manner.

In this way, it relieves the color mask from having to accurately regress each pixel

value. Only the pixels relevant to the expression change are carefully estimated,

the rest are just noise. For example, the attention is clearly obviating background

pixels allowing to directly copy them from the original image. This is a key

ingredient to later being able to handle images in the wild (see Section 6.5).

the use of the attention mask allows applying the transformation only

on the cropped face, and put it back onto the original image without producing

any artifact.

the attention mechanism not only learns to focus

on specific areas of the face but also allows merging the original and generated

image background. This allows our approach to be easily applied to images in the

wild while still obtaining high resolution images.

We crop the face with a face detector and we apply the algorithm to that specific face and that we merge it back in the image.

Note how the attention mask allows for a smooth

and unnoticeable merge between the entire frame and the generated faces.

Additionally, we embed an attention model within the network which allows focusing only on those regions of the image relevant for every speci\_c expression.

By doing this, we can easily process images in the wild, with distracting backgrounds and illumination artifacts.