Inducing Optimal Attribute Representations for Conditional GANs (Supplementary Material)

Binod Bhattarai[1] and Tae-Kyun Kim[1,2]

[1] Imperial College London, UK
[2] KAIST, Daejeon, South Korea
{b.bhattarai,tk.kim}@imperial.ac.uk

1 Additional Results

We present our additional results and insights of our work.

Simultaneous Multiple Attribute Manipulation: One of the key characteristics of our method in comparison to existing arts is to make simultaneous manipulation of frequently co-occurring auxiliary attributes to give a natural translation of the target attribute. Fig. 1 shows few more qualitative comparison of our method and its counter-part (Attgan on Diff-mode). As mentioned on the main paper, cGANs trained on Diff-mode outperforms the Std-mode. In the Fig. 1, the first, the second, and the third columns show the input image, synthetic image from the counter-part method, and synthetic image from our method respectively. First **three** rows of the Figure 1 show the outcomes when the target attribute is *female*. In all these three cases, *wearing lipstick* is clearly visible in our case whereas the counter-part method failed to do so. $P(\texttt{Wearing lipstick | male= 0.01}) => P(\texttt{Wearing lipstick | female = 0.99})$ from the co-occurrence matrix. Similarly, cheekbones are higher in all of the synthetic images from our method. Again from the co-occurrence Table, the probability of a female having *high cheekbones* is nearly 72%. Also in the second row, the *beard* is thinned a lot by our method but the counter-part method could not manage to do so. Similarly, in the third row, the intensity of colour of the hair is more *brown* than original image. Face is wrinkled, hairs removed leaving few grey hairs behind by our method when we set the target attribute as *bald*. These characterstics support the natural process of getting bald. Whereas the counter-part method just removes the hair. We can observe these differences on the fourth row the Table. In the fifth row, our method made the synthetic image look *younger* when we apply *black hair* on an *old* man with *grey hair*. In the final row, *heavy make up* turning skin to pale can be observed when we set the target attribute as *blonde hair*. The above mentioned pair of attributes are highly co-occurring attributes. We have captured such information on co-occurrence matrix and this information has been quite helpful to make such a realistic translation to target. Such meaningful translation is ultimately helping us to obtain the high target attributes recognition rate (please check experimental results on main paper).

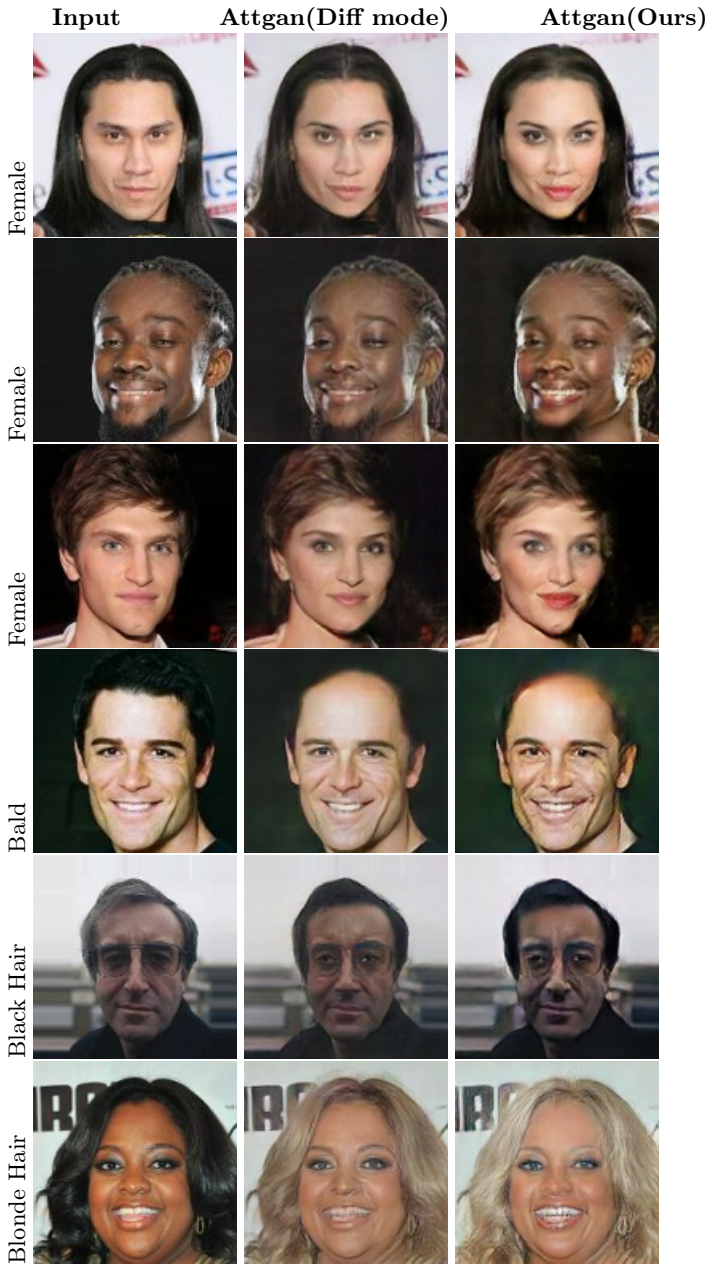

Fig. 1: Qualitative comparison on simultaneous multi-attribute manipulations. First, Second and Third column show the input image, synthetic image from the compared method (Attgan with its default conditioning on Diff mode), and Attgan with our approach respectively. The order of the target attributes are: *female*, first three rows, followed by *bald*, *black hair*, and *blonde hair*

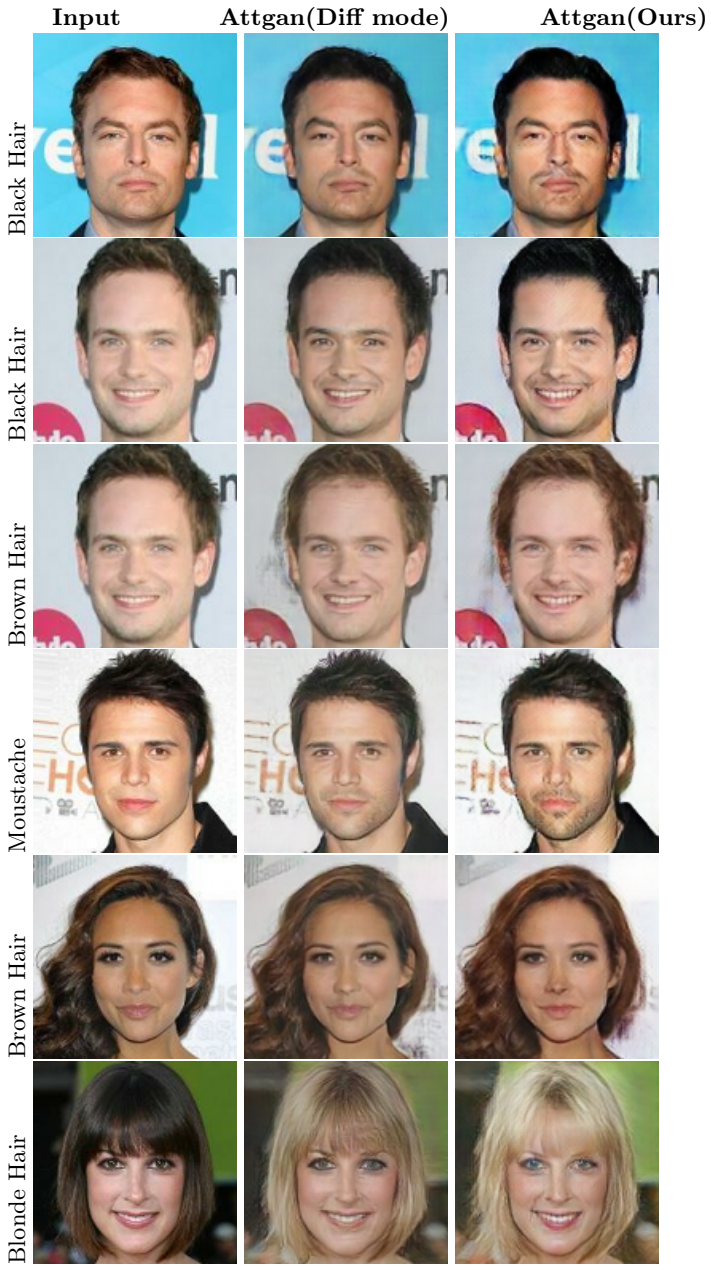

Fig. 2: Qualitative comparison on single attribute translation: First, Second and third column show the input, synthesized image from Attgan with its default conditioning on **Diff** mode, and Attgan trained with our approach, respectively. The order of target attributes are (from top to bottom): *black hair, black hair, brown hair, moustache, brown hair, blonde hair, brown hair*

Single Attribute Manipulation: In addition to doing simultaneous multiple attribute manipulation for natural translation of attributes, our method retains the target attributes in more distinct manner with less artefacts and better contrast. Fig. 2 compares the synthetic images from our method and the counter part method. From the Fig. 2, we can clearly see that our method has preserved the target attributes with higher intensities (clearer target hair colours: black, brown and blonde, and thicker moustache) and better contrasts (please compare the background too) in comparison to the counter-part method. Our method retains the co-related attributes from the source and also does not arbitarly manipulate other un-related attributes. This is demonstrated by the last three examples. In the examples, our method is able to retain *lipsticks* and *heavy make up* while setting the target attributes to *brown hair*, *blonde hair* and *brown hair* respectively. Whereas, the counter-part fainted the *lipsticks* and *make up*.

Additional Qualitative Results: Fig. 4 shows the few other synthetic examples from our method. From the Fig., we see different sets of multiple attributes manipulated when target attribute is *bald* (wrinkle on face, grey hair), *female* (lip sticks, long hair, arched eye brows), *old* (wrinkles and grey hair) and *bangs* (comparatively younger). And these sets of attributes are highly co-occurring. Thus, our method optimises generator to do simultaneous alternation of attributes for realistic translation to target domain.

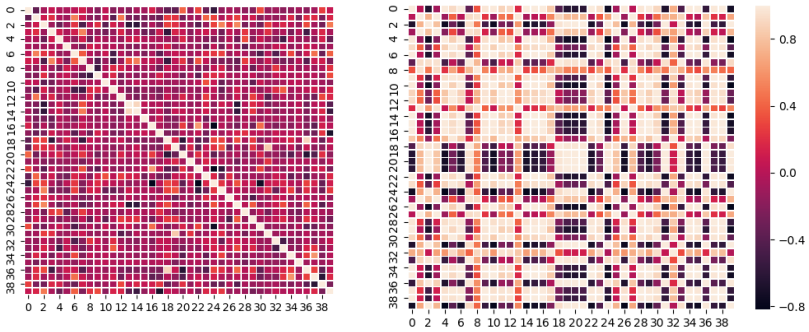

Fig. 3: Heat-map showing the cosine similarity between the initial (left) and final nodes representations by GCN (right)

Comparison with Existing Arts: Fig. 5 shows visual comparison of synthetic images with existing arts. From the Fig., we can observe that our method is able to do various auxiliary attribute manipulation depending on the target attributes, e.g. *bald* (wrinkle on face), *blonde* (heavy make up), *female* (high cheek bones, lipsticks, arched eye brows), and also exhibits clearer target attributes. Whereas amongst the existing arts, Stargan on Diff-mode and FaderNet changes auxiliary

attributes when target attribute is *female* (lipsticks). Stgan on Diff-mode and Attgan on Diff-mode manages to do so only mildly. Such multiple attribute manipulation has helped existing methods to obtain better TARR on such categories. Please Refer Tab. 2 on the main paper for empirical performance. From the Tab., FaderNet has 48.3% accuracy on male vs female although the mean TARR is 29.8%. Similarly, Stargan on Diff mode obtains +14.5% performance on male vs female above mean TARR. This suggests simultaneously manipulating auxiliary along with main attribute is beneficial to obtain high TARR.

Attributes Indices and Relation between Induced Representations: Tab. 1 shows the indices of the attributes. Fig. 3 shows the co-sine similarity between the representations of the attributes before and after GCN. Referring these two to compare the relationship between attributes, we can clearly see some of the categorical level meaningful positive and negative co-relations. For the attribute 39 (*Young*), attributes with positive co-relations are 2 (*attractive*), $18-21$ (*heavy makeup, high cheekbones, male, mouth slightly open*), 24 (*no beard*), 31 (*smiling*), 33 (*wavy hair*), 36 (Wearing Lipstick). Whereas, 4 (*bald*), 6(*big lips*), 14 (*double chin*), 15 (*eyeglasses*) are negatively co-related. Such relative relations are not distinct before the GCN (Please ref. Fig. 3, top). Similarly, 4 (*bald*) is highly un-corelated to *young* (39) which entails its positive co-relation with negation of *young* i'e *old*. Such induction are constrained by co-occurrence matrix we computed from the training examples on CelebA (See Figure 1 on the main paper) and also most of the co-relations are clearly visible on our synthetic images.

Index	Attribute	Index	Attribute
0	5 O'Clock Shadow	20	Male
1	Arched Eyebrows	21	Mouth Slightly Open
2	Attractive	22	Mustache
3	Bags Under Eyes	23	Narrow Eyes
4	Bald	24	No Beard
5	Bangs	25	Oval Face
6	Big Lips	26	Pale Skin
7	Big Nose	27	Pointy Nose
8	Black Hair	28	Receding Hairline
9	Blond Hair	29	Rosy Cheeks
10	Blurry	30	Sideburns
11	Brown Hair	31	Smiling
12	Bushy Eyebrows	32	Straight Hair
13	Chubby	33	Wavy Hair
14	Double Chin	34	Wearing Earrings
15	Eyeglasses	35	Wearing Hat
16	Goatee	36	Wearing Lipstick
17	Gray Hair	37	Wearing Necklace
18	Heavy Makeup	38	Wearing Necktie
19	High Cheekbones	39	Young

Table 1: Indices of Attributes

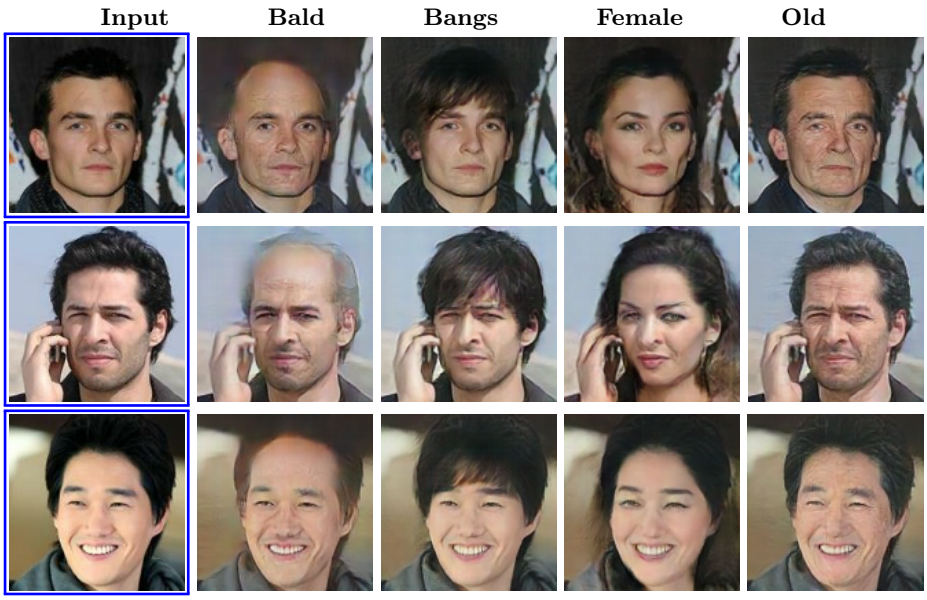

Fig. 4: Additional Qualitative results on multi-attribute manipulation after applying our method on Attgan-diff architecture. Blue boxed images are input images and rests of the images are synthetic. The order of target attributes are: *bald, bangs, female, old*

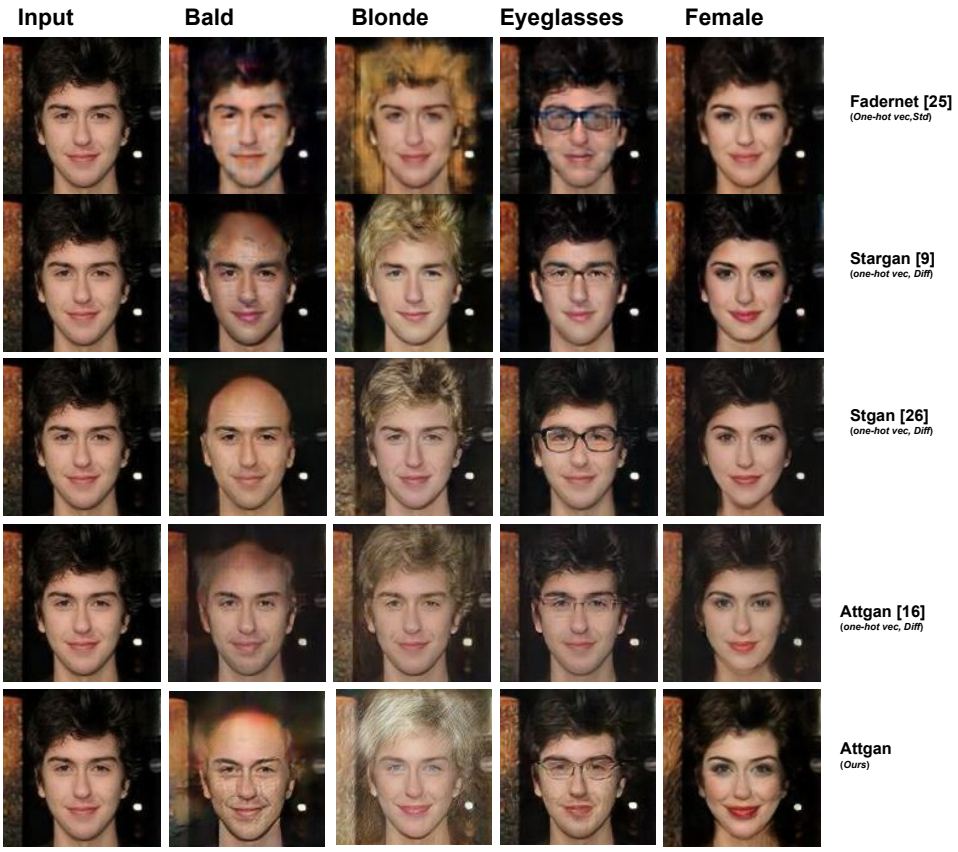

Fig. 5: Qualitative comparison of synthetic images with the existing arts for face attributes editing