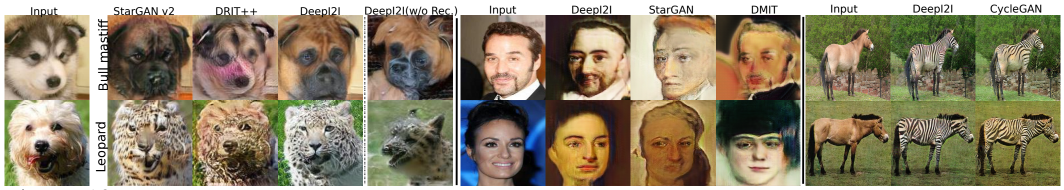
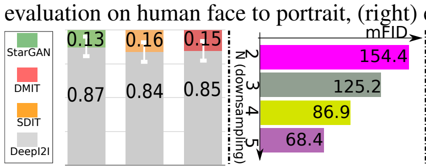


Figure 19: (left) Comparison with StarGAN v2, DRIT++, and ablation of reconstruction loss, (middle) evaluation on human face to portrait, (right) evaluation on horse to zebra translation (zoom for close inspection).



Datasets / Method	Animal faces (*710/per class*) RC↑	FC↑	mKID×100↓	mFID↓
DRIT++	35.4	33.1	14.1	138.7
StarGAN v2	39.7	40.7	6.45	85.8
DeepI2I	49.8	55.4	4.93	68.4

Enc.	Gen.	Dis.	mFID ↓
×	×	×	80.7
×	√	×	276.4
×	×	√	266.3
√	√	√	68.4

Figure 20: (a)User study, (b)study of the number downsampling, (c)related frameworks, (d)ablation of transfer learning. Note En.: encoder, Gen.: generator, Dis: discriminator

We thank the reviewers for their feedback: the paper proposes a *sound* (R1), *novel* (R4) method with *novel/new architecture* (R2/R3), obtaining *superior results* (R1, R3). It is the *first paper using pre-trained GANs for I2I initialization* (R4). We will comment the many requested experiments more completely in any final version. We will improve related work with mentioned papers.

R1 **R1.1 Discriminator I2I methods:** Figure 1(left) depicts a generative model (loosely based on BigGAN) which has not been applied to I2I before. We show that our target-label conditioning is more scalable (see also lines 98-99 in the main paper), while conditioning by one-hot vector does not scale well to many-class I2I. **R1.2 Related frameworks:** We report the result of both *DRIT++* and *StarGAN v2* in Figs. 19(left) and 20(c). We outperform them on all 4 metrics. Note that both these methods do not address transfer learning for I2I. **R1.3 Results of adaptor:** The *without adaptor* setting does have skip connections for layers 3-6 but does not have adaptor layers. The partial adaptor setting only considers a single connection. This shows that the hierarchical connections are crucial for good performance. **R1.4 Human face to portrait[29]:** we show the generated images (Fig. 19 (columns 6-9)), and obtain the FID/KID for these methods: DeepI2I/StarGAN/DMIT: (160.3/8.8) / (189.4/9.7) / (194/9.6), indicating that DeepI2I has a slight advantage.

R2 **R2.1 Evaluation metrics:** We conduct a user study and ask subjects to select results that are *more realistic given the target label, and have the same pose as the input image*. We apply pairwise comparisons (forced choice) with 26 users (30 pairs image/user). Fig. 20(a) shows that DeepI2I considerably outperforms the other methods. **R2.2 Downsampling layers:** In Fig. 20(b) we explore the downsampling layers: more downsampling layers results in better performance. **R2.3 Relation of contributions:** BigGAN-like architectures have not been explored for I2I (contr. 1). We argue and show that this helps to scale to many-class I2I problems. Directly training such architectures results in unsatisfactory results for small domains (see Tables 1 and Suppl. Mat. Sec. C). However, when combined with a pre-trained GAN (contr. 2), we obtain state-of-the-art results for many-class I2I problems, even those with fewer images per class. **R2.4 Ablation transfer learning:** We performed additional ablation of partial transfer learning in Fig. 20(d) (see also Suppl. Mat. Sec. F). In case of partial transfer networks suffer from mode collapse leading to unsatisfactory results.

R3 **R3.1 Low resolution:** we also trained our model on styleGAN with image size 256*256, and get high quality results (see Suppl. Mat. Sec. E). **R3.2 Translation of structure:** We agree with the reviewer that it would be nice to be able to measure the success of the structure translation. However, currently no evaluation metrics exist. Therefore, we propose a user study (see **R2.1**). **R3.3 Reconstruction loss:** we removed the reconstruction loss during training, and find the structure of both input (Fig. 19 first column) and output (Fig. 19 fifth column) is different, indicting that the structure information will be lost without reconstruction loss. The same loss also appears in [32][55]. **R3.4 Self-contradiction in transfer learning motivation:** The point we wanted to make here is subtle. It is not clear how to train a universal I2I network on ImageNet, since many classes cannot be translated to each other. Instead, we show the pretrained GAN features are useful for I2I. **R3.5 Horse2zebra** We performed the experiment and visualize results in Fig. 19 (right). We obtain a better FID score, i.e., (DeepI2I:CycleGAN): (63.2:77.2).

R4 **R4.1 Ablation transfer learning:** We refer to **R2.3**. **R4.2 DeepI2I (scratch and w(w/o) adaptor):** we report result in this setting, and experimentally find that the mFID is 186.4 without adaptor, and less than DeepI2I (80.7 in Tab.1) . **R4.3 Choice of w_l:** We normalize features of both Φ_l and Λ_l and used w_l=0.1 in all experiments. We found this to slightly faster convergence than w_l=1. We did not experiment with other settings. Ideally, this parameter would be optimized with cross-validation. **R4.4 Reconstruction loss:** The reconstruction loss helps align the two domains. We refer to **R3.4**.