Learning image-to-image translation using paired and unpaired training samples

Soumya Tripathy, Juho Kannala, and Esa Rahtu

Tampere University of Technology

Aalto University

Problem

Deep Image to Image translation networks are either trained with two aligned image pairs or with sets of unaligned images from two domains. These approaches have limitations such as:

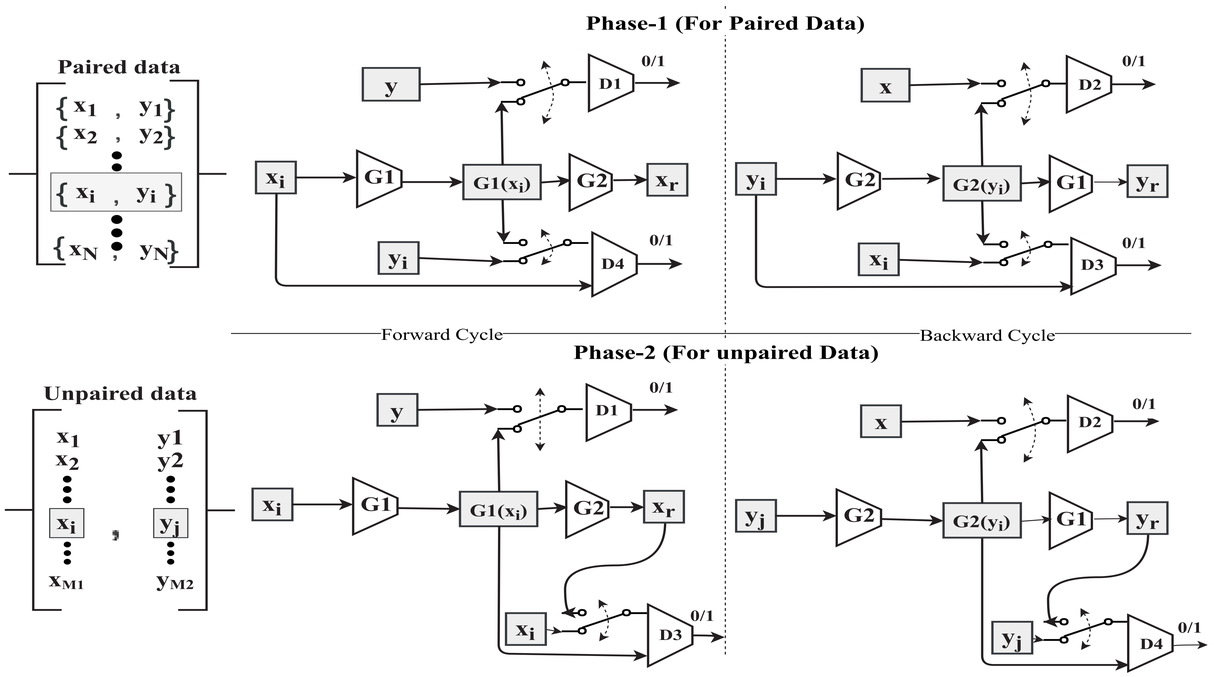
1. Image pairs provide strong supervision but for many tasks, paired data will not be available (zebra<->horse) or expensive to collect (season transfer).
2. Unaligned image pairs are easy to collect but the trained models generate low quality results due to the lack of supervision.
3. Inefficient to utilize all the available data (paired/unpaired) for a particular task.

Contributions

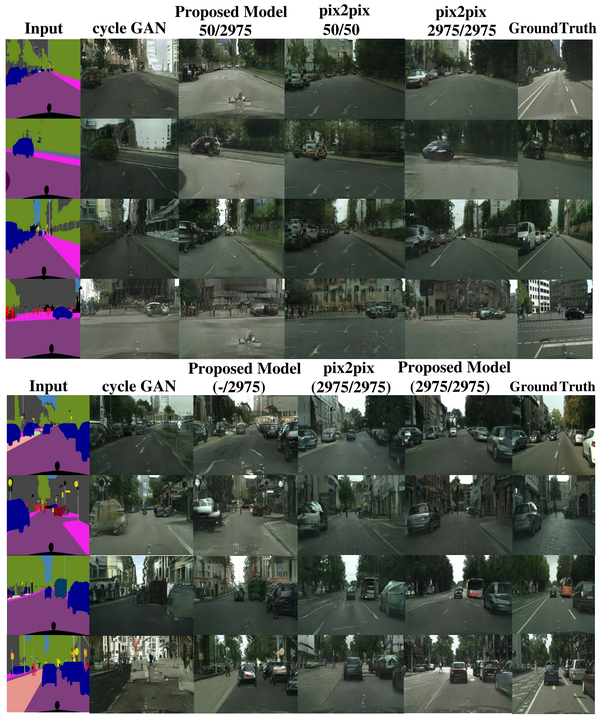
The list of contributions of our paper are:

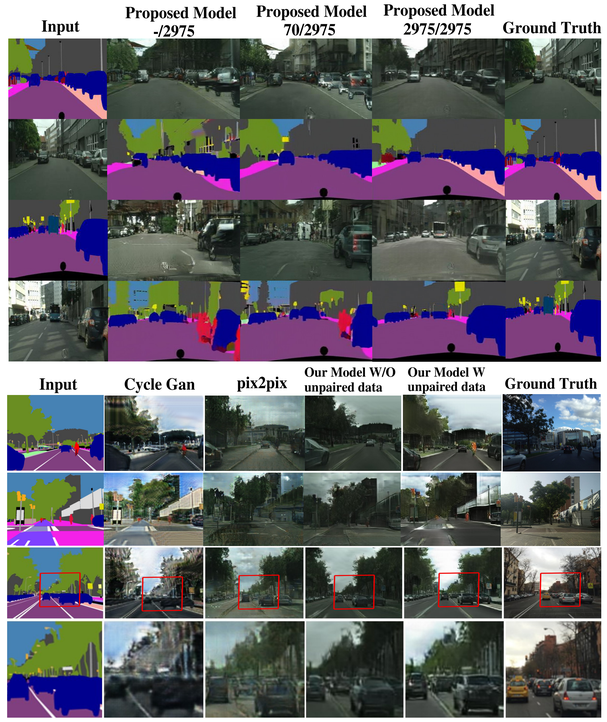
1. **We propose a new general purpose image-to-image translation model that can utilize both paired and unpaired training examples to achieve better results.**
2. A new adversarial cycle consistency loss that improves the translation results also in purely unpaired setup.
3. A combination of structural and adversarial loss improving learning with paired training samples.
4. Improving translation results by using training data from related external datasets.

Proposed Method

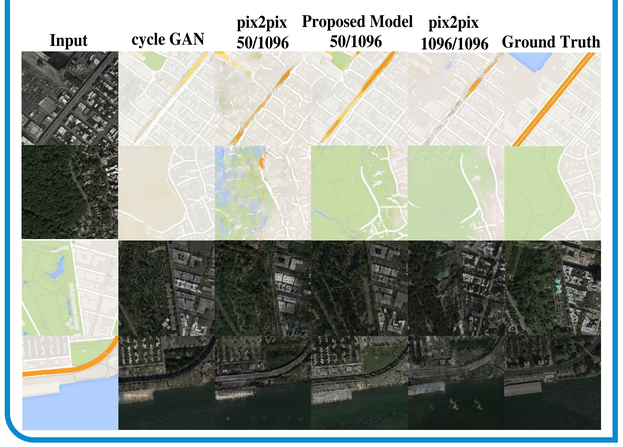


Qualitative Results





Qualitative Results



Quantitative Results

Models	paired/ total	label→photo Pixel Acc.	label→photo Mean Acc.	label→photo Mean IU	photo→label Pixel Acc.	photo→label Mean Acc.	photo→label Mean IU
Cycle GAN [1]	-/2975	0.47	0.12	0.08	0.47	0.12	0.09
Our Model	-/2975	**0.56**	**0.16**	**0.12**	**0.53**	**0.17**	**0.12**
pix2pix [2]	30/30	0.58	0.18	0.12	0.65	0.20	0.15
Our Model	30/2975	**0.66**	**0.19**	**0.14**	**0.66**	0.20	0.15
pix2pix [2]	50/50	0.60	0.17	0.12	**0.68**	**0.22**	**0.16**
Our Model	50/2975	**0.65**	**0.19**	**0.13**	0.67	0.20	0.15
pix2pix [2]	70/70	0.62	0.18	0.13	**0.70**	0.21	**0.17**
Our Model	70/2975	**0.67**	**0.19**	**0.14**	0.68	0.21	0.16
pix2pix [2]	2975/2975	0.68	0.21	0.15	0.75	0.28	0.21
Proposed	2975/2975	0.68	0.21	0.15	0.74	0.29	0.22
Proposed (VGG)	2975/2975	**0.70**	0.21	**0.16**	**0.76**	**0.32**	**0.24**

References

[1] J. Y. Zhu, T. Park, P. Isola, A. A. Efros. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks In *ICCV '17*

[2] P. Isola, J. Y. Zhu, T. Zhou, A. A. Efros. Image-to-Image Translation with Conditional Adversarial Networks. In *CVPR '17*

Conclusion

Our architecture consists of two generators and four discriminators out of which two operate in dual role depending on type of the training instance. Our model can utilise all training data irrespective of its type. The results indicated both qualitative and quantitative improvements, even in the case where training data solely consists of paired or unpaired samples.

Source Code

Try our model and let us know your experience at:
soumya.tripathy@tut.fi