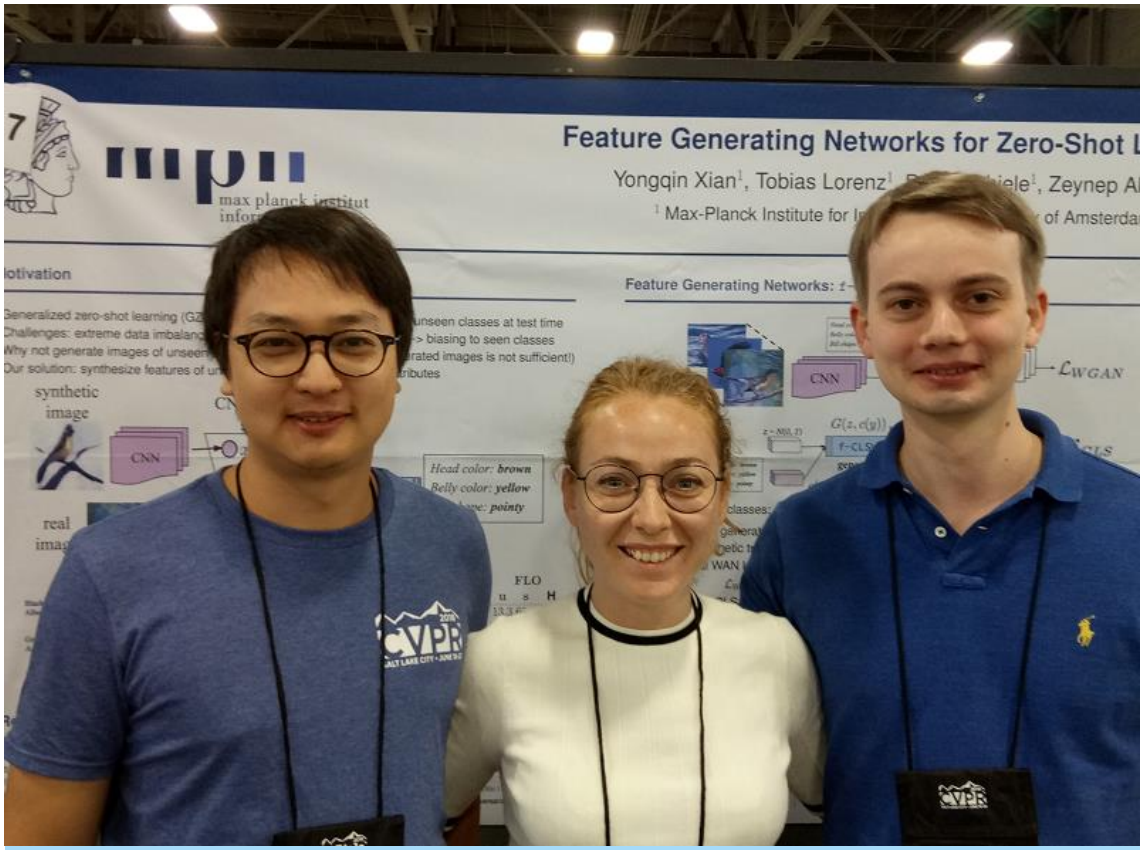

From left: Yongqin Xian, Zeynep Akata and Tobias Lorenz

Yongqin Xian is a PhD student at Max Planck Institute for Informatics, with Bernt Schiele and Zeynep Akata. We spoke to him before his poster session yesterday..

Yongqin tells us that in this work, they are trying to tackle the generalized **zero-shot learning problem**. They want to train their classifier to predict both seen and unseen classes at test time. This is different from conventional zero-shot learning, because in conventional zero-shot learning, the classifier only predicts unseen classes.

This problem is very challenging, he says, because it suffers from extreme data imbalance between seen and unseen classes. There is a lot of seen classes data, but there is no unseen classes data at all. This makes the classifier bias to seen classes. That's why in this work, he proposes to generate synthetic data of unseen classes to fix this data imbalance. However, instead of generating synthetic images of unseen classes like most **GAN** papers do, he proposes to directly generate **synthetic CNN image features** of unseen classes, conditioned on all class-level attributes.

The challenge in doing that is that they need to guarantee that the generative features are good enough to train on supervised classifiers. If it's unstable to train, they need to propose **GAN architecture** that is stable and can generate features with good quality.

… you can use this model to generate more synthetic features of those classes for which you have very few examples.

Yongqin explains what algorithmic techniques they use to solve this: "*We proposed to use Wasserstein-GAN to stabilise the GAN training. To enforce the discriminative ability of the generated features we propose a classification loss which enforces the generated features can be correctly classified by a pre-trained classifier.*"

Zeynep adds that they make use of the fact that the data that is surrounding us is inherently multimodal. If you have text that accompanies images, and you don't have enough labelled images but you have text that you can use to associate different sets of classes, then you can use this model to generate more synthetic features of those classes for which you have very few examples.

She tells us: "*We show the capability of this model on **ImageNet**. ImageNet is one of the largest scale datasets that is available to us. It generalizes to the cases when we don't have any label training data from some of the classes, just because our model is able to associate different classes in a conditional generative adversarial net framework.*"

In terms of next steps, Zeynep says they have thought about making use of the existence of text better. At the moment, they are generating image features that correspond to text sentences or attributes, but from looking at it the other way around, they can also generate text for images. They could expand this framework to do explanations, for example, explaining scene understanding and the semantic image content.