时迎琰

[shiyingyan12@qq.com](mailto:shiyingyan12@qq.com)

January 21, 2019

Brain Chip Research Center, Fudan University

develop ideas and perform original research

1. Problem Formulation
2. Solution Expression
3. Solution Execution and Evaluation

钻研：把一个思路做到极致的能力更重要

FER -> Image-to-Image -> StarGAN -> Loss

**Motivation：**获取一个人在不同姿态下的各种表情费时费力，采集麻烦，non-trivial，目前只有两个这种数据集（Multi-PIE和BU-3DFE，对于这两个数据集，有则是paired，没有则是unpaired）。表情与姿态耦合，最终的目标是为了获得更鲁棒的分类器（喂了更多样的数据即不同姿态的表情）， This paper presents DR-GAN for pose-invariant face recognition and face synthesis.

Pose variation is another common and intractable problem in unconstrained settings.

A series of GAN-based deep models were proposed for frontal-view synthesis and report promising performances. (FF-GAN, TP-GAN and DR-GAN) (source: Deep Facial Expression Recognition: A survey, 3.1.3 Face normalization: Pose normalization)

Pose coupled with expression

Facial images synthesis and pose-invariant facial expression recognition

Our model can automatically generate face images with different expressions under arbitrary poses to enlarge and enrich the training set for FER.

With one or multiple face images as the input, DR-GAN can produce an identity representation that is both discriminative and generative, i.e. the representation demonstrates superior PIFR performance, and can synthesize identity-preserving faces at target poses specified by the pose code.

Although tremendous strides have been made in facial expression recognition (FER), recognizing facial expressions in the non-frontal views are open challenge due to the limited access to large scale training data with various poses.

A majority of the proposed methods were evaluated with constrained frontal FER, and their performance degenerates when dealing with cases of non-frontal FER. To address this issue, with great efficiency and robustness.

inferior/superior performance

1.Diversity

2.Quality

3.Controllability

**思路：**利用**风格迁移**的手段生成同一个人的各种表情，即利用大量的自然场景下的表情**风格迁移**到正向表情，从而生成不同姿态下的表情。Transform the expression of a given face image to a target one without affecting the identity properties.

Different expressions under arbitrary poses

Simultaneous facial image synthesis and facial expression recognition in the wild

Neural style transfer （unpaired）

1. 实验室环境下采集到的正脸摆拍的表情（数据集1）
2. 自然场景下不同姿态的真实表情（数据集2）

变量：identity, expression, pose

Discrete / continuous: attribute value

**意义（说圆）：**扩充数据集，训练更加鲁棒的分类器，甚至可以利用不同角度的表情生成“3D模型”，用作face recognition的前处理环节（转正+无表情化，face frontalization and transformed into neutral expression）

Facial expression synthesis: synthesizing photo-realistic facial expression images has been of great value for both academic and industrial communities, and has been widely applied in facial data augmentation and face recognition. (ACM MM 2018, Geometry Guided Adversarial Facial Expression Synthesis)，最终的目标还是促进人脸表情识别的发展

研究工作的意义可以借鉴ACM MM 2018的两篇paper，写作可参考

Facial expression recognition (FER) introduction (可参考CVPR18 Joint Pose and Expression Modeling for Facial Expression Recognition): the non-frontal or the in-the-wild facial expression recognition problem is largely unexplored.

Photo-realistic face images with new identities can be generated for data augmentation, which is found to be useful to train an improved expression classifier.

Despite recent advances in face recognition using deep learning, severe accuracy drops are observed for image pose variations in unconstrained environments. Learning pose-invariant features in one-solution, but needs expensively labeled large-scale data and carefully designed feature learning algorithms. In this work, we focus on frontalizing faces in the wild under various head poses, including extreme profile view. (FF-GAN) But even representations learned with state-of-the-art CNNs suffer for profile views, with severe accuracy drops reported in recent studies. We present ablation studies to analyze the effect of each module and several qualitative results to visualize the quality of our frontalization.

Finely calibrated facial expression

To obtain a set of 68 landmarks points which depict components of human faces, such as eyes, eyebrows, nose, mouth and face out-contour. These landmarks were then linked by piece-wise linear lines of one pixel width.

Despite remarkable advances in image synthesis research, existing works often fail in manipulating images under the context of large geometric transformations.

Which can be used in various applications, including face editing, as well as 3D face reconstruction and classification of facial expression, identity and pose.

Most of the facial expression recognition systems developed so far require the subject facing the camera directly without significant head movement.

**Related work:** Facial expression recognition, Generative adversarial networks, image-to-image translation.

Generating multi-view images from a single-view input is an interesting problem with broad applications in vision, graphics, and robotics. Yet, it is a challenging problem since 1) computers need to “imagine” what a given object would look like after a 3D rotation is applied; and 2) the multi-view generations should preserve the same “identity”.

**评估：**生成表情的真实性photo-realistic，分类器的训练（原始的数据，生成的数据，微调，数据集交叉验证）

Evaluation metrics: 第三方人眼观察，Amazon Mechanical Turk, 利用人脸识别模型或人脸表情识别模型给出量化的分数

Quantitative and qualitative evaluations on two challenging in-the-wild datasets demonstrate that the proposed model performs favorably against state-of-the-art methods.

Quantitative comparisons against several prior methods demonstrate the superiority of our approach.

The facial image generation process over different iterations. ( 1,000, 6,000, 20,000, 40,000, 80,000 iterations)

We randomly select a facial image from the test set. The generated facial images with different expressions (each column) and poses (each row) are shown.

Lacking fine details and tending to be blurry.

We show that the synthesized face images have high perceptual quality, which can be used to improve the performance of an expression classifier.

This validates the synthetic face images have high perceptual quality.

Comparison of expression recognition accuracy with different numbers of synthesized images.

0, 3K, 6K, 30K, 60K

Quantitative evaluation: (StarGAN 5.5) compute the classification error of a facial expression classifier on the synthesized images. We trained a facial expression classifier on the RaFD dataset (90%/10% splitting for training and test sets) using a ResNet-18 architecture, resulting in a near-perfect accuracy of 99.55%. We then trained each of image translation models using the same training set and performed image translation on the same, unseen test set. Finally, we classified the expression of these translated images using the above-mentioned classifier.

Discuss the model’s limitations and failure cases.

The evaluation protocol includes frontal-frontal (FF) and frontal-profile (FP) face verification, each having 10 folders with 350 same-person pairs and 350 different-person pairs.

**预处理：**根据不同头部偏转角度对数据集2进行分类（水平偏角，竖直倾角，度数正负5, 15, 25, 35）

Lib face detection with 68 landmarks crop out the faces, and resize them as 256x256. For the failed images, we manually crop the faces from them.

<https://github.com/ShiqiYu/libfacedetection/>

Dlib C++ Library. face detection with 68 landmarks to crop out and align the faces.

<http://dlib.net/>.

To stabilize the training process, we design the network for the GAN based on the techniques in the CycleGAN. Specifically, this network contains two stride-2 convolutions, 9 residual blocks, and two fractionally-stride convolutions with stride 1/2. For the discriminator network, we use a 70x70 PatchGAN, which is adopted to classify whether the 70x70 overlapping image patches are real or fake.

In preliminary experiments, we also tried replacing … with … , although we did not observe improved performance.

**实现模型：** StarGAN等相关生成模型，自己增添trick

Aiming at generating photo-realistic images with high-quality local details.

Fitting data distribution

In the CycleGAN, the cycle-consistency is proposed mainly for image generation, but is agnostic to any particular task. Different from the CycleGAN, our model explicitly incorporates a task-specific classifier to enforce the global attribute-level information.

**施加约束：**表情分类器保证生成的表情，人脸验证VGGFace保证Identity，CycleGAN损失，pixel重建损失，可以涉及表情、姿态编码

身份约束应该使用高层语义信息loss，所以用face recognition的顶层卷积层的特征图的差

Identity preservation或者用于在训练过程中施加约束，或者用来检验所生成的图片的效果

**Identity Preserving Loss**（ACM MM 2018, Geometry Guided Adversarial Facial Expression Synthesis）(使用的网络源自 A lightened CNN for deep face recognition. arXiv:2015. A Light CNN for Deep Face Representation with Noisy Labels. IEEE Trans. Information Forensics and Security (2018))

|—— The discriminator on attributes disentangling, *Datt* can help the generator G learn the disentangling representation from the facial images to change the poses and expressions but retain the identity, which is useful for our FER task, because when we generate new facial images, we just want to modify the facial expression or pose of the input but without compromising the person’s identity.

To further preserve the face identity between x and x’, a pre-trained discriminative deep face model is leveraged to enforce the similarity in the feature space: we use the activations at the conv1\_2, conv2\_2, conv3\_2, conv4\_2 and conv5\_2 layer of the VGG face model. (ExprNet)

**Adversarial loss (discriminative loss)**: preserve the structural consistency of global attributes

**Pixel loss**: preserve the structural consistency of local pixels (content-similarity loss: attempts to ensure the output face sharing the expression, pose, and identity representation with the input facial image during training. L1 norm)

**Cycle-consistent Loss**

**Expression attribute loss**: Classifier C*exp* is a task-specific loss. *In the case of generation*, it can used to penalize the generator loss, which is helpful for improving the performance of the original generator G. And *in the case of classification*, it tries to classify the expression. We use a typical softmax cross-entropy loss for the classifier. We adopt the VGGNet-19 network as the classifier *Cexp*. Our classifier is trained with the GAN in an end-to-end framework.

TV(.) denotes the **total variation** which is effective in removing the ghosting artifacts. Sequentially updating the network by those losses, we could finally learn the pose-invariant FER model. (CVPR 2018, Joint Pose and Expression Modeling for Facial Expression Recognition)

We also impose a total variation regularization Ltv on the reconstructed image to reduce spike artifacts. (Understanding deep image representations by inverting them. CVPR 2015)（ExprGAN）

Conditional Expression loss: While reducing the image adversarial loss, the generator must also reduce the error produced by the AUs regression head on top of D. In this way, G not only learns to render realistic samples but also learns to satisfy the target facial expression encoded by ytarget. This loss is defined with two components: an AUs regression loss with fake images used to optimize G, and an AUs regression loss of real images used to learn the regression head on top of D. (GANimation)

**试验：**cross-dataset experiments, subject-independent 在原始数据集上训练，在另一数据集上测试。将原始数据集利用“此生成技术“扩增后训练，然后在同样的数据集上测试。其他的实验方法还有：在同一数据集上训练及测试，在另一数据集上微调或者不微调进而测试。

比较不同的分类器效果（multiSVM，AlexNet，VGGNet-16, VGGNet-19, ResNet38, ResNet50, ResNet101）(CycleAT model, ACM MM 2018)

To demonstrate the effectiveness of the proposed model, we conduct extensive experiments.

In this section, we show experimental results of our model for **facial images synthesis** and **pose-invariant facial expression recognition**. For the former task, we show qualitative results of the generated facial images under different poses and expressions. For the latter one, we quantitatively evaluate the expression recognition performance using the generated and original facial images.

We also compare our method with the models trained by different number of generated images. Given the original N images, we can randomly choose 0xN, 1xN, 5xN, 10xN, 20xN images from the generated facial images. (CVPR18 Joint Pose and Expression Modeling for Facial Expression Recognition)

Overall, our method outperforms all existing methods with a 1.88% to 7.68% improvement in terms of the FER accuracy. This may attribute to the generated images, which can help learn discriminative features to better deal with the nonlinear facial texture warping caused by poses and individual difference.

face recognition 也可以做实验的

FER on multiple head pose angles

验证时，利用face profiling approach designed to generate images with yaw angles ranging from -90 to 90，High-fidelity pose and expression normalization for face recognition in the wild, CVPR 2015

**Implementation details：**

利用facial landmarks对pose进行分类，以及裁剪出人脸

Mini-batch size of 100 through SGD

ADAM(α=0.0002， β1=0.5)

After about 50 epochs, plausible generated faces can be obtained.

Normalizing the input may make the training process converge faster.

Note that we will not use the batch normalization for E and G because it blurs personal features and makes output faces drift far away from inputs in testing. However, the batch normalization will make the framework more stable if it is applied on *Dimg*.

All intermediate layers of each block use the ReLU activation function.

Incremental training (ExprGAN): We develop an incremental training strategy to train the model on a relative small dataset without the rigid requirement of paired samples. We find in our experiments that stage-wise training is crucial to learn the desired model on the small dataset.

At the beginning of training, the reconstruction loss harms the overall process since the generation is far from frontalized, so the reconstruction loss operates on a poor set of correspondences, Thus, the weight for reconstruction loss should be set in accordance to the training stage. To reduce block artifacts, we use a spatial total variation loss to encourage smoothness in the generator output. (FF-GAN)

对合成图像的pose，identity 和expression进行评估，参考Load Balanced GANs for Multi-View Face Image Synthesis.

Compelling perceptual results

**Similarity：**与文献1的目的相似，但数据集不同，使用的模型不同，domain transfer inspired by 文献6，unified multi-domain image-to-image translation (StarGAN)

A **domain** refers to a group of images that share some latent semantic features in common, which are denoted as domain **labels**. The values of different domain labels can be either binary, like male and female for gender, or categorial such as black, blonde, and brown for hair color. The Auxiliary Classifiers GANs (ACGAN) where D is enhanced with an auxiliary classifier that learns to infer the most appropriate label for any real or fake sample. The label vector conveys semantic implications.

Future direction: towards large pose, illumination.

**人脸表情生成 相关文献：**

**(FER参考1， Image-to-image和GAN参考StarGAN)**

1. **Joint Pose and Expression Modeling for Facial Expression Recognition, CVPR 2018**

将表情和姿态disentangle，分别编码，结构是conditional adversarial autoencoder， CAAE (FER的introduction可参考) 输入的人脸得到identity code然后再加上pose和expression code

借鉴

Cexp（VGGNet-19）

Content-similarity loss施加强约束 生成的与原始的L1norm

1. Age Progression/Regression by Conditional Adversarial Autoencoder, CVPR 2017

流形学习，人脸老化退龄。结构式CAAE。 输入的人脸得到personality，然后加上age的code。

1. Facial Expression Recognition by De-expression Residue Learning, CVPR 2018

将expressive 生成 neutral

1. ExprGAN: Facial Expression Editing with Controllable Expression Intensity, AAAI 2018

编码连续控制多种表情的强度

1. GANimation: Anatomically-aware Facial Animation from a Single Image, ECCV 2018

解剖学面部变换：编辑Action Unit，实现面部动作的连续过程，不涉及姿态

1. Facial Expression in the Wild: A Cycle-consistent Adversarial Attention Transfer Approach, ACM MM 2018

将lab-controlled生成in-the-wild，利用现成的不同姿态的数据集BU-3DFE（得花钱$300）

1. Geometry Guided Adversarial Facial Expression Synthesis, ACM MM 2018

expressive和neutral的表情相互生成，涉及编码

1. StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation, CVPR 2018
2. Towards Large-Pose Face Frontalization in the wild. ICCV 2017

3D Morphable Model conditioned Face Frontalization Generative Adversarial Network, termed as FF-GAN, is proposed to generate neutral head pose face images.侧脸生成正脸，我的思路正好相反，CASIA-Net约束身份特征不变

1. Disentangled Representation Learning GAN for Pose-Invariant Face Recognition, CVPR 2017

CAAE的结构，与文献1的思路一样，identity与pose与noise拼接编码在生成同一人物的不同姿态的人脸

强调轮廓、形状的变化shape， G的功能设计

In conventional GANs, Goodfellow suggest to alternate between k (usually k = 1) steps of optimizing D and one step of optimizing G. This helps D to maintain near-optimal solution as long as G changes slowly. However, in DR-GAN, D has strong supervisions thanks to the class labels. Thus, in later iterations, when D is close to the optimal solution, we update G more frequently than D, e.g. 4 steps for optimizing G and 1 for D.

1. Beyond Face Rotation: Global and local perception GAN for photorealistic and identity preserving frontal view synthesis, CVPR 2017

identity preserving loss采用the activations of the last two layers of the Light CNN

1. Multi-channel Pose-aware Convolutional Neural Networks for Multi-view Facial Expression Recognition, FG 2018

Introduction详细介绍了Pose影响下的表情识别，可参考

1. In-the-Wild Facial Expression Recognition in Extreme Poses, arXiv:2019

Instead of avoiding and overcoming the head pose impacts, we do in the opposite approach to detect the head pose and do the expression recognition under pose awareness. 文中有head pose detection and class grouping using 68 facial landmarks

1. Unsupervised Person Image Synthesis in Arbitrary Poses. CVPR 2018 Spotlight

Conditional Pose Loss (learns how to generate samples consistent with the desired pose), 施加impose姿态标签热图的L2损失，参考pose标签的处理，test samples选择有ground truth的，方便定量评估

1. Deep Identity-aware Transfer of Facial Attributes, arXiv: 1610.05586

identity loss, 这个loss限制的是属性转换前后的图片中，人脸的身份不会丢失。对于人脸的身份信息属于高层的语义信息，文章认为并不能从图片的像素角度来定义，因而选择了卷积层的feature map来定义，采用的是VGG网络的第4层和第5层转换前后图片的feature map的平方差作为身份损失

1. Conditional Expression Synthesis with Face Parsing Transformation，ACM MM 2018

人脸表情合成的意义可参考，非端到端的训练，监督学习，两阶段的模型，产生Face Parsing Map，以此作为guidance 来指导新表情产生，identity loss采用Light CNN的特征图，Expression Preserving Loss采用一个表情分类器的输出作为约束

1. High-Fidelity Pose and Expression Normalization for Face Recognition in the wild, CVPR 2015

3D Morphable Model

1. A Light CNN for Deep Face Representation with Noisy Labels,

IEEE Transactions on Information Forensics and Security 2018, PyTorch Implementation:

<https://github.com/AlfredXiangWu/LightCNN>

1. Adversarial Training in Affective Computing and Sentiment Analysis: Recent Advances and Perspectives, arXiv:1809.08927

背景意义可参考，学位论文时用dissertation，adversarial training的发展脉络和局限，作用效果，affective computing and sentiment analysis —> facial expression synthesis and recognition ——> multi-domain image-to-image translation ——> Task-oriented GAN designed for a given task serving their own specific interests ——> expression, pose and intensity these three nonlinearly coupled factors.

1. DyadGAN: Generating Facial Expressions in Dyadic Interactions, CVPRW 2017

Expression feature vector即one-hot expression label，再加上 100-dim noise label，可以生成sketch，监督学习，和目标人脸的landmarks比较

1. Emotional Classification with Data Augmentation Using Generative Adversarial Networks,

Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD) 2018

Learning with imbalanced emotion datasets, data augmentation can be grouped into two main types: (a) geometric transformation which is relatively generic and computationally cheap; (b) task-specific or guided-augmentation methods which are able to generate synthetic samples given specific labels.

对FER样本不均衡的说明，并只针对少量样本进行GAN扩充

利用GAN进行数据增强的必要性，与平常的扩充技术的不同，对抗训练

数据集的使用及实验设置

RAFDB 训练集上增强学习，而在测试集上验证准确度，并与baseline(不用数据增强训练得到的分类器)比较

利用生成的图片训练10K steps，利用真实图片微调 10K steps

在FER2013上预训练，分别在SFEW和JAFFE上微调，再相应地进行测试

To overcome the highly imbalanced distribution

1. Geometry-contrastive generative adversarial network for facial expression transfer, arXiv: 2018, NIPS 2017 rejected，附有review和rebuttal

Contrastive learning，将68点映射到latent space保留表情信息，在和identity space拼接生成同一人物的特定表情，two-stage training，facial geometry embedding network E, image generation G/D

1. Pose Guided Person Image Generation, NIPS 2017

监督学习，two stage，先生成coarse image，然后在细化sharper图像 The use of difference maps speeds up the convergence of model training since the model focuses on learning the missing appearance details instead of synthesizing the target image from scratch. In particular, the training already starts from a reasonable result.

1. SingleGAN: Image-to-image translation by a single generator network using multiple generative adversarial learning, arXiv: 2018

图像翻译的解析可参考，one-to-many translation，many-to-many translation，one-to-one translation with multi-modal mapping

1. Toward Multimodal Image-to-image Translation, NIPS 2017

z, a low-dimensional latent space, which encapsulates the ambiguous aspects of the output mode which are not present in the input image.

1. Multimodal Unsupervised Image-to-image translation, ECCV 2018

We assume that the image representation can be decomposed into a content code that is domain-invariant, and a style code that captures domain-specific properties. To translate an image to another domain, we recombine its content code with a random style code sampled from the style space of the target domain.

1. GANGAN: Geometry-Aware Generative Adversarial Network, CVPR 2018

下面的Figure 7展示了模型效果，可以生成不同pose的人脸，不同程度的微笑，还可以更改面部纹理。模型效果是和我最像的。

1. In-the-wild Facial Expression Recognition in Extreme Poses, ICGIP 2017

针对 facial landmarks的典型处理，The samples are separated into 5 pose classes, in which pose 3 is the set of frontal faces and pose 1,2 for turning right and pose 4,5 for turning left. During the experiment, because all the photos have flipping photos, which means pose 1,2 are much similar to pose 4,5 in a mirror flipping, it will be enough for us to just implement the experiment on pose 1,2,3. All the samples are randomly separated into training and test sets, shown Table 2。 We only consider several left or right head poses, but in fact the head pose has many more poses like turn up or down and the synthesized pose of left-right and up-down [36]. Considering those effects, there would be more pose classes that need to define. The classes of Neutral and Happy occupies most the samples. If use all the samples to train the model, the model is likely to predict most sample into neutral and happy. The samples need to balance. After training the model with several iterations. a balancing sample training set should be prepared and used for continuing training. In SVM training, after training with all sample, we select the hard samples between the board line of each class pair and retrain the SVM model. This will improve the ability the recognition for fewer-samples classes.

1. Dual Generator Generative Adversarial Networks for Multi-Domain Image-to-Image Translation, arXiv: 201901

X->Y和Y->X是两个不同的G，jointly learning both the translation and reconstruction tasks with the same generator requires the sharing of all parameters, which increases the optimization complexity and reduces the generalization ability, thus leading to unsatisfactory generation performance. The dual generators, allowing for different network structures and different-level parameter sharing, are designed for the translation and the reconstruction tasks.

1. Emotion-Preserving Representation Learning via Generative Adversarial Network for Multi-view Facial Expression Recognition, 监督学习

Ablation study: To demonstrate the contribution of the loss function proposed in this paper to the final expression recognition accuracy, we perform an ablation study to evaluate the model accuracies by incrementally adding the loss terms.

1. CerfGAN: A Compact, Effective, Robust, and Fast Model for Unsupervised Mutli-domain Image-to-Image Translation, arXiv: 201805

通用的图像翻译的框架，对StarGAN的系统分析，Multi-Class Discriminator, Encoded by MCD where the input of the decoder is the feature maps of MCD

1. A Unified Feature Disentangler for Multi-Domain Image Translation and Manipulation, NIPS 2018

Encoder from pixel space to feature space, Generator from feature space to pixel space, discriminator in feature space to eliminate the domain-specific information from the representation, a discriminator in pixel space to classify the domain.

1. Identity-Adaptive Facial Expression Recognition Through Expression Regeneration Using Conditional Generative Adversarial Networks, FG 2018

The upper part is aimed to generate six basic facial expression images of the same subject for any query image using six cGANs, and each of them is designed to generate one expression respectively. 监督学习The lower part is the facial expression recognition module. A pre-trained CNN is first fine-tuned on the database, and then the last fully connected layer is used as features for both the query image and regenerated images. The query image is labeled as one of the six basic expressions based on a minimum distance in feature space. Another choice for content loss is perceptual similarity that measure high-level perceptual differences between images, based on a loss network, e.g. VGG network.

1. **Triple consistency loss for pairing distributions in GAN-based face synthesis, CVPR 2018**

GANnotation is the first method that can generate faces with a target pose and expression simultaneously. 一步生成和两步生成的目标图像应该一样的

需要提供landmarks来指示生成的人脸，故而D没有属性分类的功能

1. Every Smile is Unique: Landmark-Guided Diverse Smile Generation, CVPR 2018

We automatically extract the landmark image from the face image and encode it using a standard VAE into a compact embedding.

1. Soft-gated Warping-GAN for Pose-Guided Person Image Synthesis, NIPS 2018

监督学习，parsing to indicate regional-level segmentation

1. **An adversarial Neuro-Tensorial Approach For Learning Disentangled Representations, IJCV 2019**

Disentangling the latent factors of variations in the visual data, an end-to-end trained auto-encoder. Given a single in-the-wild image, our network learns disentangled representations for pose, illumination, expression and identity. Using these representations, we are able to manipulate the image and edit the pose or expression. 无监督张量分解，发表在顶刊上，很漂亮，很深奥，比我的工作好太多!

1. PortraitGAN for Simultaneous Emotion and Modality Manipulation，AAAI 2019

针对手机自拍，在改变不同艺术风格的同时还改变表情，facial landmarks seen as a form of pose representation, facial expressions are represented as a vector of 2D key points with N=68, 作者先用别人的算法来generate portraits of multiple styles创建desired 数据集，我也可以将各强度的表情mapping到各侧面，进而学习expression，pose，intensity三者耦合的模式

1. Show, Attend and Translate: Unpaired Multi-Domain Image-to-image Translation with Visual Attention, arXiv: 201811

Action vector instead of label vector, two different strategies of combining the input image with the target domain information: raw and latent

1. Show, Attend, and Translate: Unsupervised Image Translation with Self-Regularization and Attention, arXiv: 201806

To further constrain the learned mapping such that it is meaningful, we argue that G should preserve visual characteristics of the input image. In other words, the output and the input need to share perceptual similarities, especially regarding the low-level features. Such features may include color, edges, shape, objects, etc. We impose this constraint with the “self-regularization” term. Only use the first 3 layers of VGG. This conforms to the intuition that we would like to preserve the low-level traits of the input during translation.

1. Dual-Agent GANs for Photorealistic and Identity Preserving Profile Face Synthesis, NIPS 2017

Pose 对 FR 的影响，可参考。 Dual-Agent用来refine the profile face images generated by Simulator

1. Load Balanced GANs for Multi-view Face Image Synthesis， arXiv: 201803

Multi-view face image synthesis针对人脸，任务最相关，介绍可参考，two-stage, 先正向化，再旋转，head pose estimation作评估，加上约束即当想要的姿态是输入图像本身时，不做操作conditional self-cycle loss

1. CR-GAN: Learning Complete Representations for Multi-view Generation, IJCAI 2018

Introduce a generation sideway to maintain the completeness of the learned embedding space. Multiview generation 意义

1. Fine-Grained Head Pose Estimation Without Key points, CVPRW2018

Carefully training. We contend that low-resolution should worsen the performance of landmark detection since estimating key points necessitates access to feature which disappear at lower resolutions.

1. Reliable Crowdsourcing and Deep Locality-Preserving Learning for Expression Recognition in the Wild, CVPR 2017

Real-world Affective Faces Database <http://www.whdeng.cn/raf/model1.html>

1. Reliable Crowdsourcing and Deep Locality-Preserving Learning for Unconstrained Facial Expression Recognition. IEEE Transactions on Image Processing 2019
2. Modular Generative Adversarial Networks, ECCV 2018

组件化/模块化的GAN，将多个属性变换变为单个属性变换，这样可以级联任意多个单步变换，分为两种任务：Image generation 和 image translation

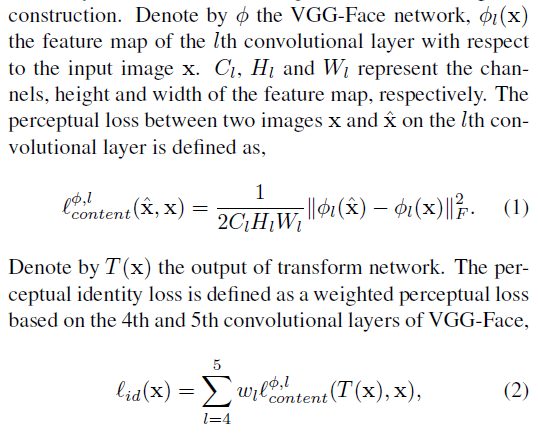
1. CR-GAN: Learning Complete Representation for Multi-view Generation, IJCAI 2018

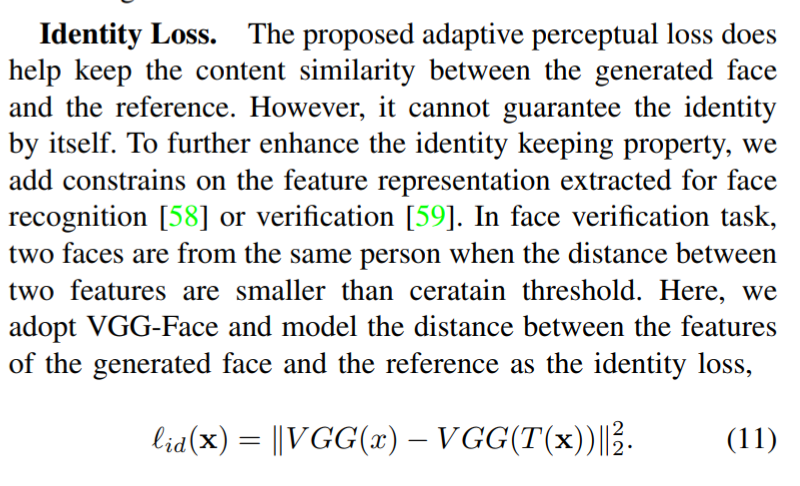
CR-GAN aims to learn complete representations in the embedding space.

1. Spatial Fusion GAN for Image Synthesis, CVPR 2019

在图片正确的位置上生成真实的物件，比如路标，警示牌。

参考文献Deep Identity-aware Transfer of Facial Attributes中identity loss定义





StarGAN的思路可借鉴，构造一个unified的GAN，用于多个域的图像翻译，可应用在多种头部姿态head pose。这篇paper认为head pose是一个attribute，而各个角度是attribute value，具有相同的attribute value的image集合构成domain。

Inspired by StarGAN, 看来要自己讲故事，发明一些新词汇用以描述自己的思路，从而区别于StarGAN。Original domain， target domain

面对的情况是创建expression + pose的新的unified version of the label as a vector

Our implementation is extensively modified from a publicly available implementation of DC-GAN.

Pose输入label直接是三种角度的值(更高层的描述)，还是landmarks的heatmap(描述更细致)？

基础：离散的姿态角度

高阶：连续的姿态角度（直接连续吧。不考虑离散的情况）

Conditioned label: how to set，3-dimension vector [pitch, yaw, roll] whose elements are normalized in the range between -1 and 1，i.e. adjusting pose angles continuously, while one-hot label (discrete angle) entails interpolation

姿态可以是连续的标签，表情是独热码标签，

若要进一步实现表情强度可控，more challenging, 需要借鉴ExprGAN中的Expression Controller Module which transforms the binary input to a continuous representation。

序列的表情过程可以用来实现表情强度

yaw是左右对称的，这点要处理，可简化运算量

Interpolation of variables: During training, we use a one-hot vector c to specify the discrete pose of the synthetic image. During testing, we could also interpolate between two neighboring pose code, to generate face images with continuous poses. This leads to smooth pose transition from one view to many views unseen to the training set.

While only discrete poses are available in training, DR-GAN can synthesize new poses by interpolating continuous pose codes.

A与B的数据集都可以用in the wild，最具普适性unconstrained setting

极致的思路：A与B均为in the wild（或为同一个数据集，或为两个数据集，如StarGAN），姿态与表情的标签均为连续值（如ExprGAN，GANimation）

[pose label, expression label]: expression label either one-hot label or randomly generated values(StarGAN)

需要注意的是：各标签分类的样本数量不均衡

区别于CAAE的框架，采用multi-domain image-to-image的思想

文章结论可提出一些general的想法

The visual appearance of objects is not only dictated by their visual texture but also depends heavily on their shape geometry.

终极目标：

产生expression，intensity，pose三种因素非线性耦合的可控图像

模型CAAE和Image-to-Image两种思路

Intensity利用视频序列数据集

key point guided

1. G2-GAN
2. Unsupervised Person Image Synthesis in Arbitrary Poses

two-stage：

1. 先从Landmarks学变换，再生成纹理（landmarks PCA减数再与原图拼接喂CNN，shape mask与texture mask）
2. 直接生成最终图像（DR-GAN pose code）

最好是端到端的训练

考虑：

1. 约束
2. 框架
3. 条件标签

impose/enforce constraints：

1. Adversarial Loss (WGAN-gp)
2. Identity loss (Light CNN)
3. Total Variation
4. Reconstruction Loss (L1)
5. conditional pose and expression code loss (auxiliary regression head on top of D)

Expression Classifier: 生成过程中用来监督，此外还可以用来分类做测试

oblation study

1. Conditional Expression Synthesis with Face Parsing Transformation, ACM MM 2018

Features:

1. Just only one dataset (RAF-DB) (15,000)
2. in-the-wild, unconstrained face (arbitrary pose)
3. unsupervised
4. synthesis (pose + expression + intensity, controllable) (yaw symmetry, reduce complexity)
5. novel GAN architecture/framework
6. extensive experiment (FER 2013, KDFE, SFEW 2.0)

样本数量：

ExprGAN:

only used Oulu-CASIA dataset 1440 images (1296 for training, 144 for testing)

GANimation:

EmotioNet 200,000 images out of one million

StarGAN:

CelebrA 20,0000 for training, 2000 for testing

RaFD 4,824 images collected from 67 participants.

Geometry-Guided GAN:

CK+ 593 sequences from 123 subjects (training and testing subsets are divided based on identity with 100 for training and 23 for testing);

Oulu-CASIA videos of 80 subjects with six typical expressions (60 subjects for training, 20 subjects for testing)

CycleGAN:

Maps <—> aerial photograph (1096),

Cityscapes <—> Photo (2975),

Horse(939) <—> Zebra(1177),

Apple(996) <—> Orange(1020),

Summer(1273)<—> Winter(854)

Age Progression:

Because the proposed approach does not require multifaces from the same subject, we simply randomly choose around 3000 images from the Morph and CACD dataset and crawl 7,670 images from the website. We divide the age into ten categories.

The final dataset consists of 10,670 face images with a uniform distribution on gender and age.

Joint pose and expression model:

Multi-PIE (6,124 for training, 1,531 for testing),

BU-3DFE (16, 800 for training, 4,200 for testing)

**Code repository:**

1. ExprGAN
2. CycleGAN
3. Pixel2pixel
4. Age progression/regression
5. Joint pose and expression model (TF)
6. GANimation
7. StarGAN
8. DR-GAN
9. TP-GAN (TF)

**人脸表情数据集的模式：**

1. 同一个人的多个表情，实验室摆拍场景，正脸 constrained frontal expression
   1. CK+ (593序列)
   2. MMI（213序列）
   3. Oulu-CASIA (480序列)
   4. JAFFE (213)
   5. TFD (112,234)
   6. KDEF (4900) （5 different yaw angles:-90, -45, 0, 45, 90）
2. 一个人只有一个表情，无限制场景下的自然表情，各种姿态(unconstrained facial expressions, varied head poses, changed illumination, large age range, different face resolutions, occlusions, and varied focus) in which the facial expressions are spontaneously displayed in real-world environment.
   1. FER 2013 (35,887) (used to test the trained classifier as well)
   2. ExpW (91,793)
   3. RAF-DB (29,672)
   4. AffectNet (450,000)
   5. SFEW 2.0 (1,400) (used to test the trained classifier to validate the effectiveness of data augmentation)

within-dataset and cross-dataset settings

To verify the generalization of the proposed method, cross-dataset experiments were carried out, as shown in Table III.

RAF-DB

KDEF

FER2013

SFEW2.0

**RAF-DB** diversity in posture, illumination, person identity, expression distribution.

强调 为了更贴近实际情况，为了获得more robust and accurate FER，特地选择Real-world Affective Faces

RAF-DB/basic/README.txt :

train\_xxxx.jpg 12271

test\_xxxx.jpg 3068

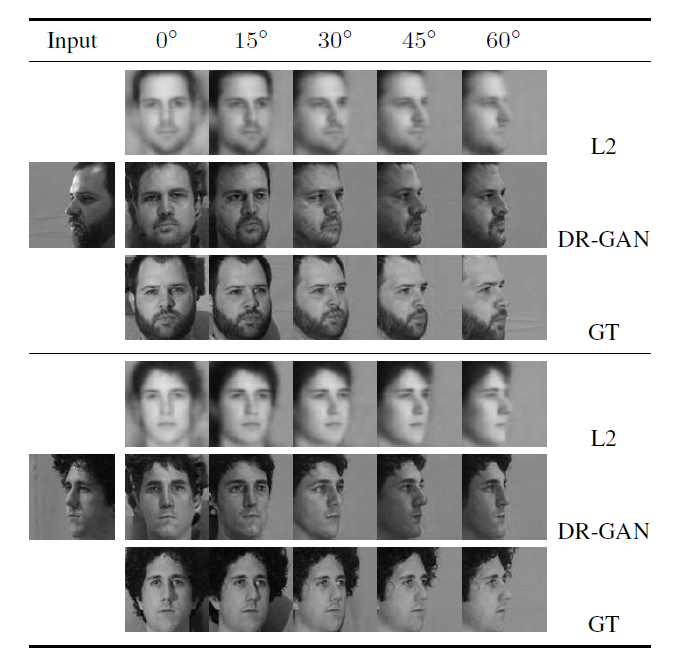
First roughly aligned using similarity transformation according to the two eye locations and the center of mouth;

Then resized to 100\*100

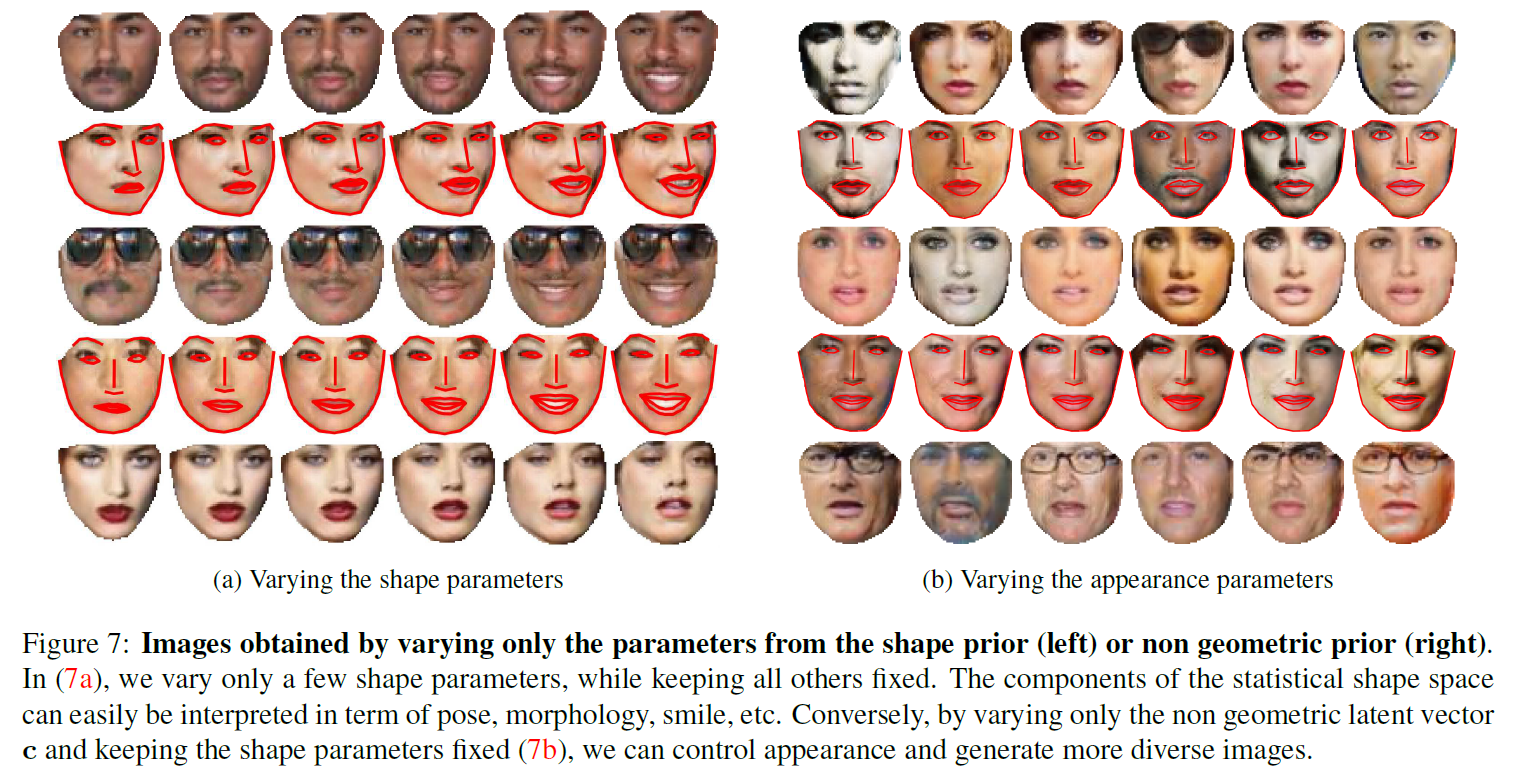
Fine-grained head pose estimation without key points, CVPRW 2018

Georgia Institute of Technology

<https://github.com/natanielruiz/deep-head-pose>



GAGAN (CVPR 2018):



处理后的SFEW 2.0 - qq\_33894186的博客 - CSDN博客

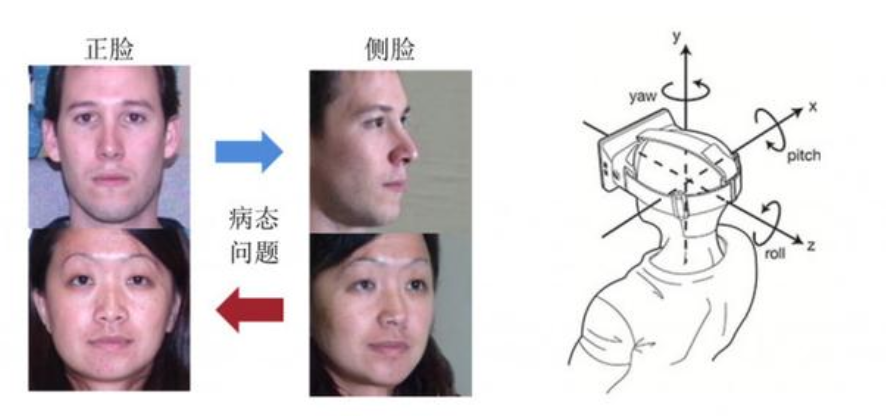
<https://blog.csdn.net/qq_33894186/article/details/80779343>

中科院自动化所赫然：大规模人脸图像编辑理论、方法及应用

<http://baijiahao.baidu.com/s?id=1599526577293416704&wfr=spider&for=pc>

生成模型，即学习联合概率密度分布，它可以从统计的角度表示数据的分布情况，能够反映同类数据本身的相似度。生成模型的主要功能有两个：一是进行密度估计，二是生成样本。生成/合成人脸时，所要的就是生成/合成的人脸和真实人脸相似。生成模型中大家比较熟悉的就是GAN，即生成对抗网络。

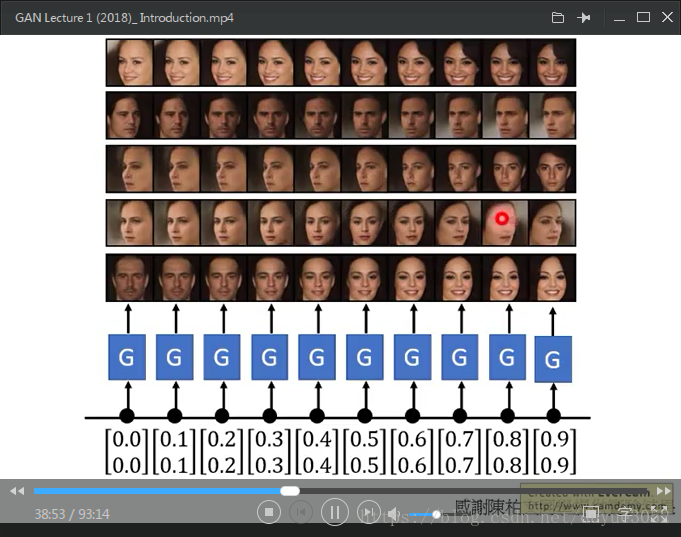
**人脸视角旋转**应用，即将归一化的人脸旋转到任意姿态。例如从一张正脸图像生成侧脸图像；或反之，从采集到的一张侧脸恢复其正脸图像，公安领域常有此需求。



视角旋转有 x、y、z 三个方向，我们目前只考虑左右偏转。如果从单张图像进行旋转的话，这需要「无中生有」，因为有些信息是没有的，所以旋转时结果存在偏差。人脸旋转有两部分研究内容，一部分是 2D 模型，一部分是 3D 模型。

我们希望机器能够对现有的图像进行自动处理，并且得到一些新的图像，而这些新图像则需要同时符合人的认知和特定的需求。该问题是当前机器学习、计算机视觉重要的研究内容之一，并且在交互娱乐、卫生医疗、公共安全等领域有着广泛的应用场景。

朱俊彦 dissertation: Learning to synthesize and manipulate natural images



对GAN中生成器的输入向量插值，就会出现头部姿态旋转的效果。

DR-GAN：Disentangled Representation Learning GAN for Pose-Invariant Face Recognition论文解读

<https://blog.csdn.net/qq_34914551/article/details/87365119>

DR-GAN代码实现记录

<https://blog.csdn.net/qq_34914551/article/details/87381230>

DR-GAN源码-CSDN下载

<https://download.csdn.net/download/qq_34914551/10959638>

kayamin/DR-GAN

<https://github.com/kayamin/DR-GAN>

HRLTY/TP-GAN

<https://github.com/HRLTY/TP-GAN>

lzhbrian/image-to-image-papers: 🦓<->🦒 🌃<->🌆

A collection of image to image papers with code (constantly updating)

<https://github.com/lzhbrian/image-to-image-papers>

Paper Notes: Cross-Domain Image Translation Based on GAN

<https://blog.csdn.net/java_n4a/article/details/79761856>

AlfredXiangWu/face\_verification\_experiment:

Original Caffe Version for LightCNN-9.

<https://github.com/AlfredXiangWu/face_verification_experiment>

Highly recommend to use PyTorch Version

<https://github.com/AlfredXiangWu/LightCNN>

**新思路：**

参考deep 3D Morphable Model conditioned Face Frontalization GAN（FF-GAN, ICCV 2017）考虑基于3DMM的pose + expression生成。

关于3D模型的生成应用 或许可以参考这篇文献的介绍NIPS2015

Weakly-supervised Disentangling with Recurrent Transformations for 3D View Synthesis

一个博士（机器学习方向）关于发论文的几点忠告 - zhinanpolang的专栏 - CSDN博客

<https://blog.csdn.net/zhinanpolang/article/details/83045414>

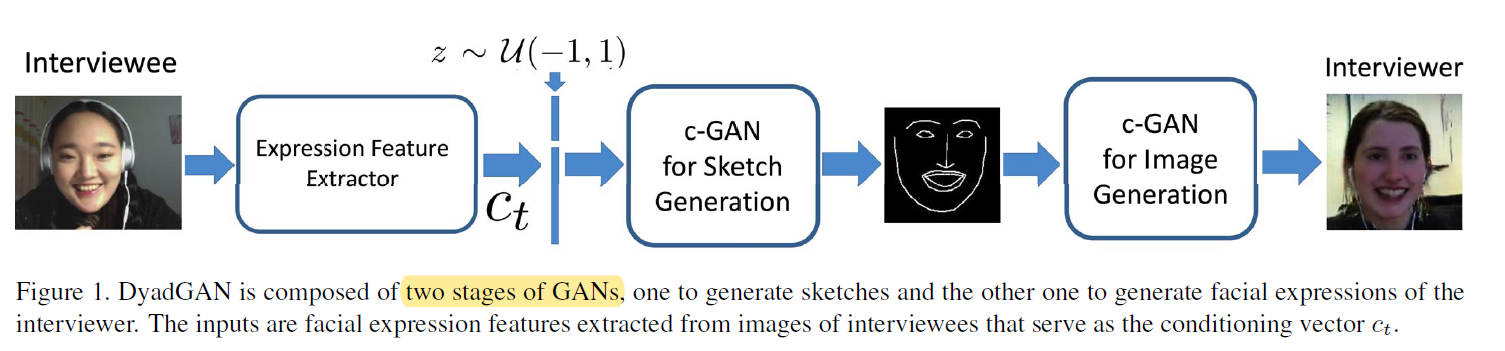
计算机视觉研究入门全指南----新手博士需要准备的资料 - maweifei的博客 - CSDN博客

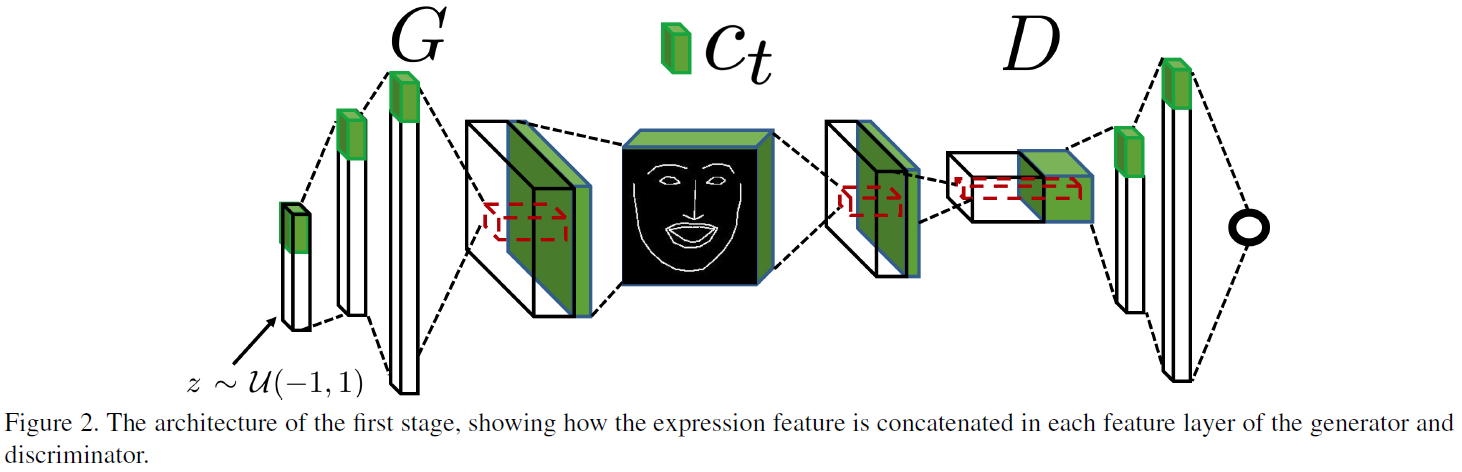
<https://blog.csdn.net/maweifei/article/details/82385917>

人脸属性分析--性别、年龄和表情识别 - 迷若烟雨的专栏 - CSDN博客

<https://blog.csdn.net/minstyrain/article/details/82257369>

DyadGAN: Generating Facial Expressions in Dyadic Interactions, CVPRW 2018





For expressive face sketch generation, two stages are both DC-GAN.

For sketch to image generation, a sketch is passed through an encoder with 8 down-sampling layers, and then a decoding composed of 8 up-sampling layers to produce an image. U-Net strategy.

