

Abstract

Recently, with the development of deep Convolutional Neural Networks (CNNs) and large-scale datasets, face recognition (FR) has made remarkable progress. However, recognizing faces in large poses and under heavy occlusion remains a vital challenge due to the unbalanced training data. Thanks to Generative Adversarial Neural networks (GANs), synthesizing photorealistic multi-view faces and unveiling heavily occluded face images becomes feasible, which significantly facilitates FR and has a wide range of applications in entertainment and art fields. This thesis provides an in-depth study of GAN-based face image synthesis and its application to FR.

The current facial image synthesizing methods have featured two main research lines, i.e., 2D-based and 3D reconstruction-based. Our works cover both of them. For 2D-based face pose editing, current methods primarily focus on modeling the identity preserving ability but are less able to preserve the image style properly, which refers to the color, brightness, saturation, etc. This thesis proposes a novel two-stage approach to solve the style in-consistency problem, where face pose manipulation is cast into pixel sampling and face inpainting. With pixels sampled directly from the input image, the face editing result faithfully keeps the identity and the image style.

For traditional 3D face reconstruction, due to the linear and low-dimensional nature of the 3D Morphable Model (3DMM), the reconstructed textures hardly capture high-frequency details, resulting in blurred textures that are far from satisfactory. Some recent 3D face reconstruction methods have also leveraged adversarial training to improve the texture quality. However, these methods either rely on scarce, non-public 3D face data or complex and costly optimization approach. This thesis proposes a high-fidelity texture generation method, which predicts the global texture of the 3D face from a single input face image. The training is based on the pseudo ground truth blended by the 3DMM and input face textures. Multiple partial UV map discriminators are leveraged to handle the imperfect artifacts in the pseudo ground truth.

In terms of face de-occlusion, we propose a Segmentation-Reconstruction-Guided face de-occlusion GAN, consisting of three parts, a 3DMM parameter regression module N_θ , a face segmentation module N_S , and an image generation module N_G . With the texture prior provided by N_θ and the occluded parts indicated by N_S , N_G can faithfully recover the missing textures. The proposed method outperforms the state-of-the-art methods quantitatively and qualitatively.

Keywords: face pose editing, 3D face reconstruction, facial image de-occlusion.

Résumé

Récemment, avec le développement des "Convolutional Neural Networks"(CNNs) et des ensembles de données à grande échelle, la reconnaissance des visages (RF) a fait des progrès remarquables. Cependant, la reconnaissance de visages dans de grandes poses et sous forte occlusion reste un défi vital en raison du déséquilibre des données d'entraînement. Grâce aux "Generative Adversarial Networks" (GANs), il est possible de synthétiser des visages multi-vues photoréalistes et de dévoiler les images de visages fortement occlus, ce qui facilite considérablement la RF et offre un large éventail d'applications dans les domaines du divertissement et de l'art. Cette thèse fournit une étude approfondie de la synthèse d'images de visages basée sur les GAN et de son application à la RF.

Les méthodes actuelles de synthèse d'images faciales présentent deux axes de recherche principaux, à savoir les méthodes basées sur la 2D et celles basées sur la reconstruction 3D. Nos travaux couvrent les deux. Pour l'édition de la pose du visage en 2D, les méthodes actuelles se concentrent principalement sur la modélisation de la capacité de préservation de l'identité mais sont moins capables de préserver correctement le style de l'image, qui fait référence à la couleur, la luminosité, la saturation, etc. Cette thèse propose une nouvelle approche en deux étapes pour résoudre le problème d'incohérence de style, où la manipulation de la pose du visage est divisée en échantillonnage de pixels et en peinture du visage. Avec des pixels échantillonnés directement à partir de l'image d'entrée, le résultat de l'édition du visage conserve fidèlement l'identité et le style de l'image.

Pour la reconstruction traditionnelle de visages en 3D, en raison de la nature linéaire et basse dimensionnelle du modèle morphable 3D (3DMM), les textures reconstruites capturent à peine les détails à haute fréquence, ce qui donne des textures floues qui sont loin d'être satisfaisantes. Certaines méthodes récentes de reconstruction de visages en 3D ont également exploité l'entraînement contradictoire pour améliorer la qualité de la texture. Cependant, ces méthodes s'appuient soit sur des données de visage 3D rares et non publiques, soit sur une approche d'optimisation

complexe et coûteuse. Cette thèse propose une méthode de génération de texture haute-fidélité, qui prédit la texture globale du visage 3D à partir d'une seule image de visage en entrée. L'apprentissage est basé sur la pseudo vérité de terrain mélangée par le 3DMM et les textures du visage d'entrée. De multiples discriminateurs de cartes UV partielles sont utilisés pour gérer les artefacts imparfaits de la pseudo-vérité terrain.

En termes de désocclusion de visage, nous proposons un GAN de désocclusion de visage basé sur segmentation et reconstruction, composé de trois parties, un module de régression des paramètres 3DMM N_θ , un module de segmentation des occlusions N_S , et un module de génération d'images N_G . Avec la texture préalable fournie par N_θ et les parties occluses indiquées par N_S , N_G peut récupérer fidèlement les textures manquantes. La méthode proposée surpasse quantitativement et qualitativement les méthodes de l'état de l'art.

Mots clés: synthèse de pose de visage, reconstruction de visage tridimensionnelle, dé-occlusion d'images faciales.