Facial Keypoint Detection with Quaternion Convolutional Neural Networks

ZHANG HUAKANG LI JIALIN YU HANG

March 4, 2022

Table of Content

Introduction

Facial Key Points (FKPs) detection is an important and challenging problem in the field of computer vision, which involves detecting FKPs like centers and corners of eyes, nose tip, etc. FKPs detection can be applied in tracking faces in images and videos, analysis of facial expressions, detection of dysmorphic facial signs for medical diagnosis, face recognition, etc.
In the past few years, advancements in FKPs detection are made by the application of deep convolutional neural network (DCNN), which is a special type of feed-forward neural network with shared weights and local connectivity. Not only in FKPs detection, but in other computer vision tasks, CNN is widely used and becomes more and more mature in recent years.

Introduction

CNN still has some drawbacks, like when dealing with the color images,general CNNs just treat the RGB three channels as three unrelated feature maps. For each kernel it just sums up the outputs corresponding to different channels and ignores the complicated interrelationship between them. We may lose important structural information of color and obtain non-optimal representation of color image.
Focusing on the problems mentioned above, we are going to propose a model using on FKPs called quaternion convolutional neural network which represents a color in the quaternion domain. Its conventional kernel, pooling layer and full connected layer will be replace with the operation of quaternion algebra.

Related Works

Facial Keypoint Detection

Facial keypoints detection is a problem of estimating the position of eyes, nose, and mouth in a facial image. This problem, also known as face alignment, has been widely studied for many years in the field of computer vision because of its relevance to various face analysis applications such as face recognition , face attribute recognition, head pose estimation, and 3D face modeling systems.

In order to develop a detector that is robust to disturbances and environmental changes, existing studies have proposed a feature extraction algorithm[1, 2] or a method that can directly model the shape of the face[3, 4]. Recently, various CNN-based regression methods have been proposed like deep convolutional network cascade for facial point detection[5].

Related Works

CNN

Convolutional neural network is one of the most successful models in many vision tasks. Since the success of LeNet[6] in digit recognition, great progresses have been made. AlexNet[7] is the first deep CNN that greatly outperforms all past models in image classification task. Then, a number of models with deep and complicated structures are proposed, such as VGG[8] and ResNet[9].A traditional convolutional neural network consists of one or several convolutional layers, followed by some fully-connected layers of neurons. Each convolution block usually produces feature maps by four steps, *e.g.*, convolution, batch normalization, non-linear activation, and pooling.

Related Works

Quaternion

Quaternion is a kind of number system extending the complex numbers which was first described by William Rowan Hamilton in 1843. A quaternion $\hat{q}$ in the quaternion domain $\mathbb{H}$, *i.e*, $q \in \mathbb{H}$, can be represented as $\hat{q} = q_0 + q_1 i + q_2 j + q_3 k$, where $q_n \in \mathbb{R}$ for $n = 0, 1, 2, 3$, and the imaginary units i, j, k obey the quaternion rules that $i^2 = j^2 = k^2 = ijk = -1$.

Related Works

Quaternion

Similar to real numbers, we can define a series of operations for quaternions:

- Addition:

$$\hat{p} + \hat{q} = (p_0 + q_0) + (p_1 + q_1)i + (p_2 + q_2)j + (p_3 + q_3)k$$

- Scalar multiplication:

$$\lambda\hat{q} = \lambda q_0 + \lambda q_1 i + \lambda q_2 j + \lambda q_3 k$$

- Element multiplication:

$$\begin{aligned}\hat{p}\hat{q} =& (p_0q_0 - p_1q_1 - p_2q_2 - p_3q_3) + (p_0q_1 + p_1q_0 + p_2q_3 - p_3q_2)i \\ &+ (p_0q_2 - p_1q_3 - p_2q_0 - p_3q_1)j + (p_0q_3 + p_1q_2 - p_2q_1 - p_3q_0)k\end{aligned}$$

- Conjugation:

$$\hat{q}^* = q_0 - q_1 i - q_2 j - q_3 k$$

These quaternion operations can be used to represent rotations in a three-dimensional space. Suppose that we want to rotate a 3D vector $\mathbf{q} = [q_1, q_2, q_3]^T$ to get new vector $\mathbf{p} = [p_1, p_2, p_3]^T$, with an angle θ and along a rotation axis $\mathbf{w} = [w_1, w_2, w_3]^T$, $w_1^2 + w_2^2 + w_3^2 = 1$. Such a rotation is equivalent to the following quaternion operation:

$$\hat{p} = \hat{w}\hat{q}\hat{w}^* \tag{1}$$

where $\hat{q} = 0 + q_1 i + q_2 j + q_3 k$, $\hat{p} = 0 + p_1 i + p_2 j + p_3 k$ and

$$\hat{w} = \cos\frac{\theta}{2} + \sin\frac{\theta}{2}(w_1 i + w_2 j + w_3 k) \tag{2}$$

Related Works

Quaternion

Since its convenience in representing rotations of three-dimensional vectors, in the field of computer vision and image processing, quaternion-based methods show its potentials in many tasks. For example, quaternion principal component analysis network[10] and quaternion-based sparse representation of color image[11] have been proven to be better in extracting more representative features for color images.

Motivation

In recent years, real-valued CNN has been widely used in the field of computer vision. CNN has a good performance in all kinds of tasks, the image classification and retrieval, target location detection, object segmentation and face recognition. In face recognition, each person's facial features vary greatly, and for the same person, due to the size, position, posture and so on, the change will be different. This is more difficult under different conditions, such as lighting, viewing Angle, occlusion, etc[12]. Convolutional neural network with deep structure can solve these problems well.

Motivation

But CNN also has some drawbacks. Its natural judgment of color images is defective[13]. In the specific case of image recognition, a good model has to efficiently encode local relations within the input features,such as between the Red, Green, and Blue channels of a single pixel.
In particular, traditional real value CNN treats pixels as three separate values. In other words, in the training process of CNN, both internal implicit relation and global implicit relation are considered at the same level.
Therefore, CNN may lose some information that depends on the internal relationship of the image.

Motivation

To solve this problem, we propose to replace CNN with QCNN, which treats a pixel as a multidimensional entity. In particular, each color pixel in a color image is representedas a quaternion, and accordingly, the image is represented as a quaternion ma-trix rather than three independent real-valued matrices. QCNN enhances the ability of the model to learn the internal relations of pixels and external relations.

Proposed method

Basic Architecture

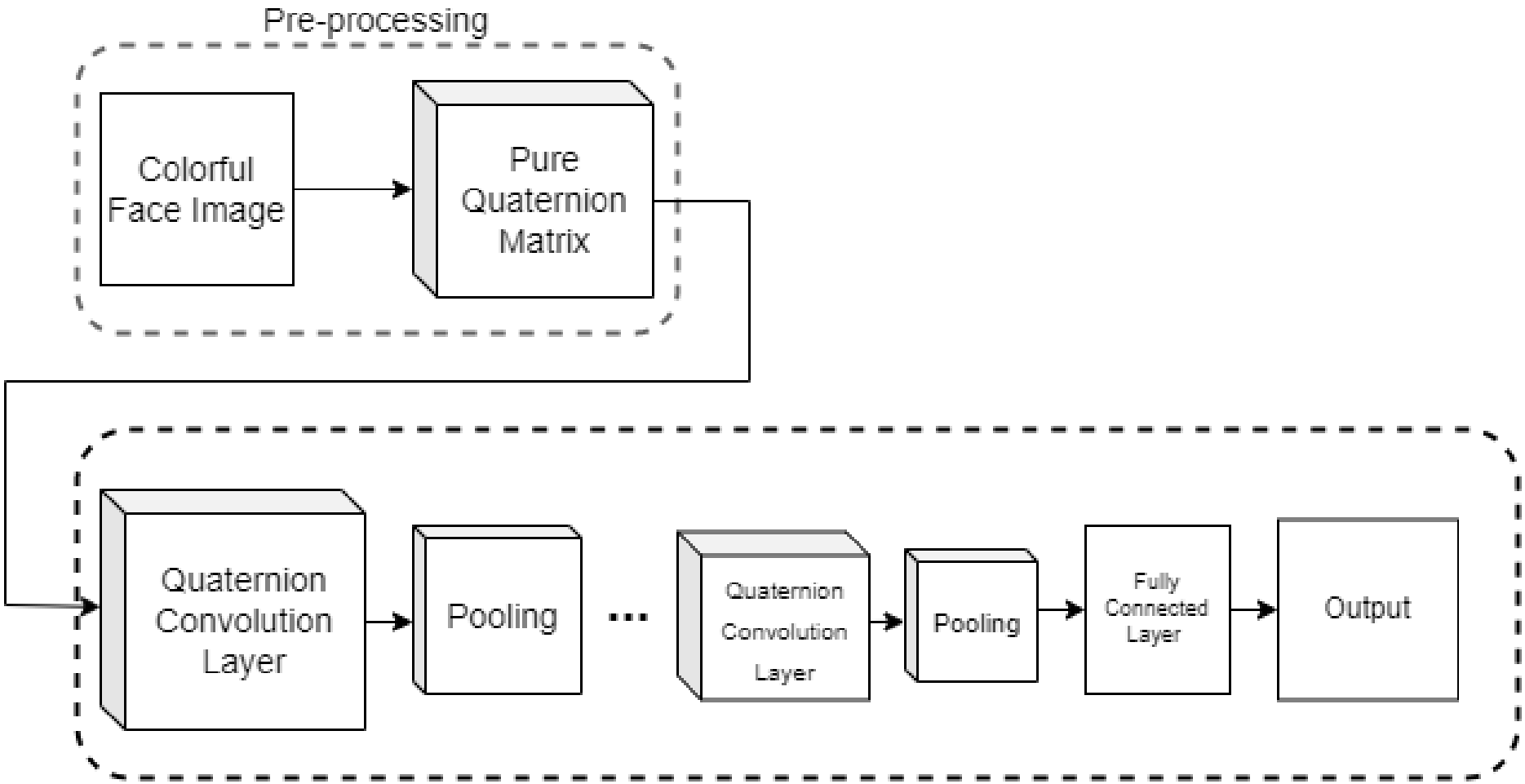


Figure: Architecture

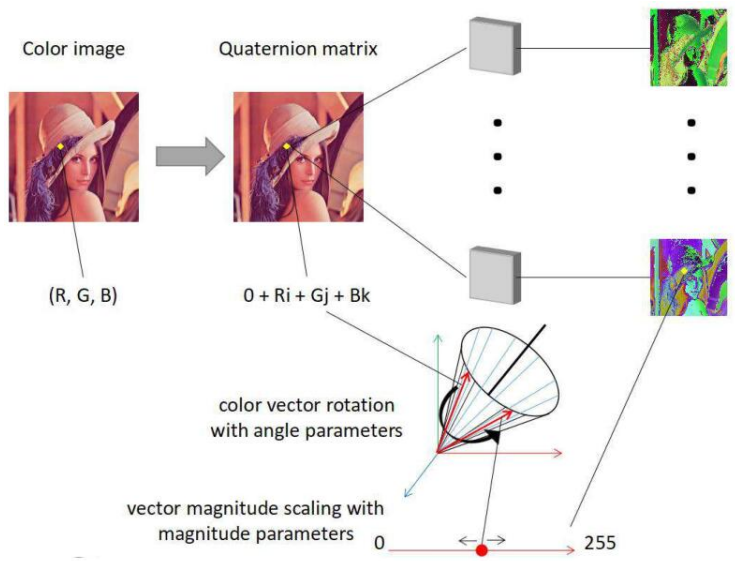
Proposed method

Pure Quaternion Matrix

Focusing on color image representation, our quaternion CNN treats a color image as a 2D pure quaternion matrix, denoted as $\hat{A} = [\hat{a}_{nn'}] \in \mathbb{H}^{N \times N}$ where N represents the size of the image. In particular, the quaternion matrix $\hat{A}$ is

$$\hat{A} = 0 + \mathbf{R}i + \mathbf{G}j + \mathbf{B}k, \tag{3}$$

where $\mathbf{R}, \mathbf{G}, \mathbf{B} \in \mathbb{R}^{N \times N}$ represent red, green and blue channels, respectively.



Proposed method

Quaternion Convolution, Pooling & Fully Connected Layer

- **Quaternion Convolution Layer**: We are going to come up with a new quaternion convolution kernel and an operation between the input and kernel based on the quaternion rotation.
  - Which axis the rotation should follow.
  - How to determine the angle of rotation.
- **Pooling Layer & Fully Connected Layer** : We are going to come up with a new pooling layer and fully connected layer based on the quaternion algebra, *e.g.*, addition, scalar multiplication and conjugation.
  - Which pooling function in the real number domain can be extended to quaternion.
  - How does the pooling operation work.
  - How to define the fully connected layer.
- **Backpropagation** is the key of training a network, which applies the chain rule to compute gradients of the parameters and updates them.

- **Backpropagation** is the key of training a network, which applies the chain rule to compute gradients of the parameters and updates them.
  - How to define the loss function.
  - How to get the backpropagation algorithm in QCNN?

Proposed method

Implement

Similar to real-valued CNN, QCNN can be accelerated using parallel computing, which means training on GPU with thousands of cores will be faster than training on CPU with dozens of cores. Thus, we are going to develop our QCNN framework based on *CUDA API* or a machine learning framework that supports GPU computing, *e.g. PyTorch*, *Tensorflow*.

Thanks

Q&A


T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, pp. 2037–2041, Dec 2006.


Z. Cao, Q. Yin, X. Tang, and J. Sun, "Face recognition with learning-based descriptor," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2707–2714, June 2010.


T. F. Cootes, "An introduction to active shape models," 2000.


T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," in *Computer Vision — ECCV'98* (H. Burkhardt and B. Neumann, eds.), (Berlin, Heidelberg), pp. 484–498, Springer Berlin Heidelberg, 1998.


Y. Sun, X. Wang, and X. Tang, "Deep convolutional network cascade for facial point detection," in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3476–3483, 2013.


Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, pp. 2278–2324, Nov 1998.


A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems* (F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, eds.), vol. 25, Curran Associates, Inc., 2012.


K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.


K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, June 2016.


R. Zeng, J. Wu, Z. Shao, Y. Chen, B. Chen, L. Senhadji, and H. Shu, "Color image classification via quaternion principal component analysis network," *Neurocomputing*, vol. 216, pp. 416–428, 2016.


L. Yu, Y. Xu, H. Xu, and H. Zhang, "Quaternion-based sparse representation of color image," in *2013 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1–7, July 2013.


S. Colaco and D. S. Han, "Facial keypoint detection with convolutional neural networks," in *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, pp. 671–674, Feb 2020.


T. Parcollet, M. Morchid, and G. Linarès, "Quaternion convolutional neural networks for heterogeneous image processing," in *IEEE ICASSP*, (Brighton, United Kingdom), May 2019.