

基于角度特征的分类网络

王立华 杜明辉 梁亚玲

华南理工大学电子与信息学院 广州 510641

(1132967661@qq.com)



摘要 卷积神经网络(Convolutional Neural Networks,CNN)在图像分类任务中的卓越表现,使得其被广泛应用于计算机视觉的各个领域。图像分类模型精度与效率的提升,除了归功于网络结构的改变外,还有很大一部分原因来自于归一化技术以及分类损失函数的改进。在人脸识别任务中,随着精度的不断提升,分类损失函数从 Softmax Loss 到 Triplet Loss,又从 L-Softmax Loss 到 Arcface Loss,度量方式从几何度量发展到角度度量。度量方式的改变实际上是特征形式的变化,即特征形式从一般特征转变为角度特征。在 Mnist 数据集上,使用角度度量损失函数训练得到的特征点呈角度分布,同时准确率比几何度量高;将角度度量方式用更直接的角度特征来表示,训练得到的同类特征点呈直线分布,准确度也比一般角度度量更高。这不禁令人思考,在 CNN 分类模型中是否可以使用角度特征来代替一般特征。在 CNN 分类模型中,其主要架构往往由多个卷积层和一个或多个全连接层组成,通过统一卷积层与全连接层的归一化操作,得到角度卷积层与角度全连接层。在普通分类网络的基础上,用角度卷积层替换卷积层,用角度全连接层替换全连接层,可以得到一个由角度特征组成的角度分类网络。在 Cifar-100 数据集上,基于 ResNet-32 构造的角度分类网络相比原分类网络,分类准确率提高了 2%,从而论证了角度特征在分类网络中的有效性。

关键词: 图像分类;卷积神经网络;归一化;损失函数;角度特征

中图法分类号 TP183

Classification Net Based on Angular Feature

WANG Li-hua, DU Ming-hui and LIANG Ya-ling

School of Electronics and Information, South China University of Technology, GuangZhou, 510641, China

Abstract The excellent performance of Convolutional Neural Networks (CNN) in image classification tasks makes CNN models widely used in various fields of computer vision. In addition to the changes in the network structure, a large part of the reason why the accuracy and efficiency of the image classification model increase year by year comes from the normalization technology and the improvement of the classification loss function. In the face recognition task, with the increasing precision, the classification loss function change from Softmax Loss to Triplet Loss, and from L-Softmax Loss to Arcface Loss, the measurement method develops from geometric measurement to angle measurement. The change of measurement mode is actually a change of feature form, and the feature form changes from general feature to angle feature. The feature points trained on the Mnist dataset using the angle metric loss function are angularly distributed, and the accuracy is higher than the geometric metric. If the angle metric is represented by more direct angular features, the feature points of the same class are linearly distributed after training, and accuracy is also higher than the general angle metric. This makes people wonder whether angle features can be used instead of general features in the CNN classification model. In the CNN classification model, the main structure is often composed of multiple convolutional layers and one or several fully connected layers. Through unifying the normalization operation of the convolutional layer and the fully connected layer, layers in model come to the angular convolutional layers and the angular fully connected layers. On the basis of the common classification network, the convolution layer is replaced by the angle convolution layer, and the full connection layer is replaced by the angle full connection layer, and then an angle classification network composed of angular features can be obtained. The accuracy of the angle classification network constructed on ResNet-32 is 2% higher than that of the original classification network on the Cifar-100 dataset. The validity of the feature in the classification network is demonstrated.

Keywords Image classification, Convolutional neural networks, Normalization, Loss function, Angular feature

到稿日期:2019-05-17 返修日期:2019-08-16 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金资助项目(61701181);广东省自然科学基金资助项目(2017A030325430);广州市科技计划项目(201707010070)

This work was supported by the National Natural Science Foundation of China (61701181), Natural Science Foundation of Guangdong Province, China (2017A030325430) and Science and Technology Program of Guangzhou, China (201707010070).

通信作者:杜明辉(ecmhdu@scut.edu.cn)

近年来,卷积神经网络在图像分类任务中的卓越表现^[1-3],使得其被广泛应用在图像检测^[4-6]、图像语义分割^[7]、图像动作检测^[8]等各个计算机视觉领域。图像分类模型精度与效率的提升,除了归功于网络结构的改变^[9-10]外,还有很大一部分原因来自于归一化技术^[11-13]和损失函数的改进^[14-20]。在人脸识别任务中,分类损失函数从 Softmax Loss 发展到 Triplet Loss^[14],又从 L-Softmax^[15]发展到 Arcface Loss^[20];随着特征归一化和权重归一化的推广应用,度量方式从几何度量发展到角度度量,精确度不断提升,在作为人脸识别任务基准之一的 Lfw 数据集^[21]上精确度甚至达到了 99.83%^[20]。受此激励,我们也开始尝试在 CNN 分类模型中使用角度特征替代数值特征,希望网络的每一层只输出角度特征,剔除模长特征。CNN 分类模型的主要架构往往由多个卷积层和一个或多个全连接层组成。Liu^[22]提出了一种卷积层归一化方式 Sphereconv,这种卷积方式可以剔除卷积核的模长信息和输入特征的模长信息,得到仅与角度相关的输出特征。而人脸识别分类损失函数在 Softmax Loss 基础上进行的一系列改进,也可以看作对全连接层的改进。Ranjan^[16]指出全连接层的输入特征模长与图片质量相关,将其归一化可以剔除无效信息;Liu^[17]也证明了归一化全连接层的权重模长可以减小训练数据不平衡产生的影响。而经过特征归一化和权重归一化的输出特征恰恰只剩下了角度信息。

本文通过统一卷积层与全连接层的归一化操作,将卷积层与全连接层改造为角度卷积层与角度全连接层,每层仅输出角度特征,得到一个由纯角度特征组成的角度分类网络;最后,在 Cifar-100 数据集上验证了所提模型的有效性。

本文的主要工作如下:1)在 Mnist 数据集上通过可视化探究角度度量的物理含义;2)提出角度全连接层;3)修改 Sphereconv 角度函数,使得角度特征能直观显示;4)将角度全连接层与角度卷积层统一起来;5)利用角度卷积层与角度全连接层构建角度分类网络;6)研究在不同角度函数与参数选择下,角度分类网络在 Cifar-100 数据集上的分类准确率。

1 背景介绍及相关工作

1.1 Softmax Loss 分类损失函数的一系列改进

Softmax Loss 其实包含了最后一个全连接层。

$$y_j = \sum_{j \in C_m} x_j * w_{i,j} \quad (1)$$

$$p_i = \frac{e^{y_i}}{e^{y_i} + \sum_{j \neq i} e^{y_j}} \quad (2)$$

其中, x_j 表示第 j 个输入特征, y_i 表示第 i 个输出特征, $w_{i,j}$ 为对应于 x_j 与 y_i 的权重, p_i 为第 i 个输出特征所对应的预测概率。而在对特征与权重进行归一化的过程中,实际上只有全连接层改变了。归一化后,有:

$$y_i = \cos(\theta_i) = \frac{\mathbf{X} \cdot \mathbf{W}_i}{|\mathbf{X}| |\mathbf{W}_i|} \quad (3)$$

其中, \mathbf{X} 和 \mathbf{W}_i 分别为 (x_1, x_2, \dots, x_m) 与 $(w_{i,1}, w_{i,2}, \dots, w_{i,m})$ 的向量表示。

在对 Softmax Loss 的角度度量进行改进的一系列版本中,我们选择由 Liu^[15]提出的 L-Softmax Loss 与由 Deng^[20]提出的 Arcface Loss 作为代表进行讨论。L-Softmax 与 Arcface Loss 对 Softmax Loss 进行改进的过程中,剔除特征模长

与权重模长后,在训练时将 Softmax 函数中的 $\cos(\theta)$ 分别用 $\cos(m\theta)$ 与 $\cos(m+\theta)$ 替代,而 y_i 保持不变,以获得良好的训练性能;在测试时,使用原本的 $\cos(\theta)$ 数值。如果将 Softmax 函数看作一个普通的指数归一化方式,那么这些修改实际上是在修改最后一个全连接层。

1.2 Sphereconv 卷积归一化方式

对于一个常规卷积操作,假设输入图片的大小为 $W \times H \times C_{in}$,其任意位置的数值为 $X(w, h, c_{in})$,卷积核参数的数量为 $K \times K \times C_{in} \times C_{out}$,其任意位置的数值为 $F(k_i, k_j, c_{in}, c_{out})$,则在忽略偏置的情况下,输出特征图中任意位置的卷积运算结果为:

$$Y(w_m, h_n, c_{out}) = \sum_{i \in K} \sum_{j \in K} \sum_{k \in C_{in}} X(w_{m+i}, h_{n+j}, c_k) * F(w_i, h_j, c_k, c_{out}) \quad (4)$$

将这个运算需要的所有特征点的值放到一个向量 \mathbf{X} 中,所有的卷积核参数的值放到一个向量 \mathbf{W} 中,在忽略偏置的情况下,就得到:

$$\mathbf{Y} = \mathbf{X} \cdot \mathbf{W} \quad (5)$$

Sphereconv 操作是分别计算 \mathbf{X} 和 \mathbf{W} 的模长,最后得到:

$$\mathbf{Y} = \cos(\theta) = \frac{\mathbf{X} \cdot \mathbf{W}}{|\mathbf{X}| |\mathbf{W}|} \quad (6)$$

即输出仅与角度 θ 有关,为角度特征。

1.3 Decoupled Networks(DCNet)

Liu^[23]提出了一种去耦合网络,通过使用不同的角度函数和模长函数的组合来改进 Sphereconv 操作,以使得性能最优。本文研究很大一部分建立在该研究的基础上,但主要的关注点不在于如何选择模长函数和角度函数,而是探究在整个网络(包括卷积层与全连接层)中使用角度特征的可行性以及角度特征选取的方式。

2 探究分析

2.1 损失函数可视化与角度度量

为了充分体现角度度量,定义全连接层中的角度特征为:

$$\cos(\theta_i) = \frac{\mathbf{X} \cdot \mathbf{W}_i}{|\mathbf{X}| |\mathbf{W}_i|} \quad (7)$$

$$y_i = -\log(\theta_i + m) \quad (8)$$

这一层的输出经过 Softmax 函数处理后,计算得到的损失被称作纯角度损失(Pure Angular Loss)。为了简单地可视化不同损失函数的分类效果,我们选用 Mnist 手写数据集这个最简单的图像分类基准数据集作为可视化实验的目标数据集。在 Mnist 手写数据集上,将 LeNet-5 作为 CNN 网络,分别使用 4 种损失函数,即 Softmax Loss, L-Softmax Loss, Arcface Loss 和 Pure Angular Loss,并将网络输出特征维数设定为 2。可视化结果见 OSID 码。

在实验过程中,我们经过细微的参数调节,针对每一个损失函数选择其最优参数。L-softmax loss 选择参数 $s=5, m=4$; Arcface Loss 选择参数 $s=5, m=0.6$; Pure Angular Loss 选择参数 $m=0.3$ 。值得注意的是,使用 L-softmax 分类的特征点有两簇重叠,这是由于 Liu^[23]提到需要在训练前期使用 Softmax Loss 使得训练收敛,而我们省略了这个步骤,故特征点簇产生了重叠。

由图 1 可得到以下信息:

1)以 Softmax Loss 作为分类损失函数得到的特征点呈现角度分布,说明 Softmax Loss 本身也是一种隐式的角度度量方式;

2)Softmax Loss, L-Softmax Loss, Arcface Loss 和 Pure Angular Loss 作为分类损失函数得到的特征点分布,随着同类特征点越来越趋于直线(最好情况为一条直线),精确率不断提高(除了 L-Softmax Loss 因为没有使用 Softmax Loss 来帮助训练,特征点簇产生重叠而影响了精确率,后续分析也忽略这个特殊情况);

3)在 Pure Angular Loss 中,特征点几乎以一条直线排列,说明 CNN 完全有能力仅通过角度信息完成分类任务。

2.2 角度度量的物理含义

从实验中可以看出,对于使用角度度量损失函数训练的模型而言,其训练过程就是一个寻找合适的向量 W 作为类向量的过程。训练时,调整各层权重,使得输出的所有特征点向各自类向量靠拢,最后模型收敛为几条直线(代表着几个类)。

特征向量的方向实际上代表着各个特征数值的不同比例。对于两个特征向量,如果它们处于同一个方向,意味着它们的各个特征数值比例一致。假设有 3 个物体,如图 1 所示,图 1(a)由一个圆与一条线段组成,图 1(b)由一个小圆与一条小线段组成,图 1(c)由一个圆与两条线段组成。图 1(a)~图 1(c)的特征向量数值 (x, y) 分别为: $(1, 1)$, $(0.5, 0.5)$, $(1, 2)$ 。特征 x 与 y 分别代表了圆与线段,以人的直观判定,图 1(a)与图 1(b)为一类,图 1(c)为另一类。对于同一类图像,往往会由于光照、尺寸等不同,导致其特征向量数值整体发生变化。因此,特征数值之间的比例关系对于类别判定而言就显得尤为重要,换句话说,特征值的不同比例就代表了不同的类。

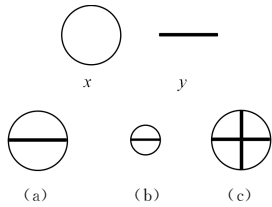


图 1 特征向量的物理含义示意图

Fig. 1 Schematic diagram of physical meaning of feature vector

2.3 误分类点的分布

从前文分析可以看到,特征向量的模长实际上代表了它所属这一类的“置信度”,模长越大,说明它的可信度越高;模长越小,说明可信度越低。从图 1 中可以清晰地看出,几种损失函数训练的分类模型的特征点交叠部分(误分类区域)都主要集中在原点附近,这也与人的直觉一致。Ranjan^[16]指出特征模长与图片质量呈正相关,对于人的直观而言,针对清晰的高质量图片,人能更准确地判断其类别,反之则越容易判定错误;而对于分类模型而言,特征模长越大,其分类的“置信度”越高,由于原点附近的特征模长很小,因此分类模型对其分类时越容易出现误分类情况。

3 算法模型

3.1 角度全连接层

根据上面的探究,定义角度全连接层的输出为:

$$y = -\log(\theta), \theta = \arcsin\left(\frac{X \cdot W}{\|X\| \|W\|}\right) \quad (9)$$

选用 Log 函数作为角度函数对角度进行数值变换,是因为我们希望角度函数具备以下特征:1)角度函数为角度的单调递减函数;2)角度函数的导函数为角度的单调递减函数。第 2)点主要是考虑物理意义。对于分类任务而言,我们希望经过网络得到的特征向量与类向量的角度均趋于 0;对于大角度而言,角度数值没有太大意义,因为都偏离了类向量;而对于小角度而言,0.2 与 0.1 的角度代表的意义是完全不同的,它们表示了不同的分类置信度,故需要将小角度以更加细粒的方式区分出来。

对于 Log 角度函数,在实验中我们发现,当输出的特征维度过小(比如这里为了可视化,将维度设为 2)时,会观察到训练不收敛的现象。为了使得训练能有效收敛,我们添加了一个小数值 m ,通常这个数值范围在 0~0.5 之间。于是:

$$y = -\log(\theta + m) \quad (10)$$

其中, m 越小,角度函数在 $x=0$ 点处的导数越大。当特征维度过小时,特征向量能表示的方向较少(如维度为 2 时,其只能表示一个平面内的方向),此时不同特征向量与类向量之间的角度很容易重合,使得输出的角度接近零。在这种情况下,当 m 数值过小(如 $m < 0.1$)时,角度函数在 $x=0$ 点处的导数将会过大,容易出现训练不收敛的情况。而 m 数值越大,其整体函数图像在 $[0, \pi)$ 范围内会越接近于线性函数,通常其取值不会超过 0.5。

与 L-Softmax, Arcface Loss 相比,角度度量的 Pure Angular Loss 有如下优势:1)物理意义明确,直接输出角度特征;2)对小角度有更好的区分度,使得在训练时,小角度特征点能继续向类向量靠拢,最终使得特征点以直线形式排列;3)训练与推理时函数形式保持一致,而不像 L-softmax 与 Arcface Loss 那样在模型训练与推理时使用不同的函数处理特征;4)特征点更趋于直线,分类精确率也更高。我们将这种直接输出角度特征的全连接层称为角度全连接层。

3.2 角度卷积层

Liu^[22]在 Sphereconv 中使用了 3 种角度函数:

$$y = 1 - \frac{2}{\pi} \theta \quad (11)$$

$$y = \cos(\theta) \quad (12)$$

$$y = \frac{1 + e^{-\frac{\theta}{2k}}}{1 - e^{-\frac{\theta}{2k}}} * \frac{1 - e^{\frac{\theta}{k} - \frac{\pi}{2k}}}{1 + e^{\frac{\theta}{k} - \frac{\pi}{2k}}} \quad (13)$$

与角度全连接层类似,我们使用 Log 函数作为角度函数,这几个角度函数的函数图像见 OSID 码。从图中可以看到,其他函数在 $x=0$ 点附近的导数较小, Cosine 和 Sigmoid 函数在 $x=0$ 点附近的导数甚至接近于 0。这使得这些函数在训练过程中,当训练的特征点运动到类向量附近时,由于其导数已经接近于 0,因此就停滞在类向量周围;而 Log 函数能很好地解决这个问题,特征点越接近类向量,就认为此时训练的方向越正确,应提供更大的导数,使得特征点继续向类向量运动。因此,相比于其他函数在小角度下的“不作为”,Log 函数能有效地体现小角度之间的差异,同时激励特征点呈直线分布,后续在 Cifar-100 上的实验验证了 Log 函数的有效性。

我们把经过角度函数处理后的 Sphereconv 称为角度卷积层,角度函数默认采用 Log 函数。

3.3 角度全连接层与角度卷积层的联系

通常情况下,我们认为全连接层是一种卷积核大小为

1×1 的特殊卷积层。而实际上,我们也可以把卷积层看作一种全连接层,卷积层相当于多个并行的全连接层,每个全连接层输入($K \times K \times C_{in}$)维的特征向量 \mathbf{X} ,与($K \times K \times C_{in}$) \times C_{out} 数量的权重 \mathbf{W} 相乘后求和,输出 C_{out} 维的特征向量 \mathbf{Y} 。于是,角度全连接层与角度卷积层的输出,即角度特征的计算公式可以统一如下:

$$\mathbf{Y} = -\log(\arccos\left(\frac{\mathbf{X} \cdot \mathbf{W}}{\|\mathbf{X}\| \|\mathbf{W}\|}\right)) \quad (14)$$

利用我们定义的角度卷积层和角度全连接层替换普通 CNN 模型,得到一种角度分类网络。

4 实验设置与实验结果

将 ResNet-32 作为基础模型,用角度卷积层和角度全连接层分别替换 ResNet-32 模型原有的卷积层与全连接层,得到我们的角度分类网络。采用 DCNet 的配置,使用 Batch Norm 归一化、正交正则化和 L2 正则化,选用 Relu 作为激活函数,设置 Batch Size 为 128,初始学习率为 0.001,使用阶梯型学习率策略,在 Cifar-100 数据集上进行图像分类实验,且分别在 epoch 为 20000,30000,37500 时,将学习率减小到上一阶段的 1/10。

4.1 角度缩放系数的探究

Ranjan^[16] 提出为了使得经过 Softmax 函数归一化的输出在训练时收敛,需要对输入 Softmax 函数的数值进行放大。于是,Softmax 函数变为:

$$p_i = \frac{e^{S \cdot y_i}}{e^{S \cdot y_i} + \sum_{j \neq i} e^{S \cdot y_j}} \quad (15)$$

其中, y_i 为上一层的输出, S 为缩放系数。事实上,这个缩放系数将底数 e 变为了 e^S 。在类别较多的情况下,即使其他的 y_j 都为 0,由于指数函数的特性,每一个 y_j 都会在分母中占据 1,使得正确类的输出过低,而增大底数能有效改善这一问题。

我们验证了不同的缩放系数 S 值的影响效果,实验结果如表 1 所列。根据实验结果,我们简单地认为在 $S > 10$ 时,模型能达到较好的收敛效果,在 $S = 30$ 时模型效果最优。因此,在后续的实验中,我们将从角度全连接层输出的数值统一放大 30 倍,再进行 Softmax 函数运算。

表 1 在不同缩放系数 S 数值下角度分类网络在 Cifar-100 上的分类准确率

Table 1 Classification accuracy of angular classification networks on Cifar-100 under different scaling factor S values

S 值	分类准确率/%
1	51.72
5	75.84
10	77.44
20	78.08
30	78.22
40	77.98

4.2 附加小数值 m 的设定

3.1 节已提到,为了防止角度函数数值过大,我们在 Log 函数中附加了一个小数值 m ,函数的完整形式如式(16)所示:

$$y = -\log(\theta + m) \quad (16)$$

对不同的 m 值进行实验,结果如表 2 所列。

表 2 不同 m 取值下角度分类网络在 Cifar-100 上的分类准确率

Table 2 Classification accuracy of angle classification networks on Cifar-100 datasets under different values of m

m 值	分类准确率/%
0	78.22
0.01	77.83
0.1	78.26

通过实验我们发现,在特征维度数量较大的情况下(这里为 384),并没有发现训练不收敛的情况,故在后续的研究中,再设置这个 m 参数,这使得我们的模型更加简洁。

4.3 对比实验

将普通卷积层、Sphereconv 和角度卷积层与普通全连接和角度全连接层进行两两组合,得到不同分类模型。各个分类模型在 Cifar-100 上的分类精确度如表 3 所列。

表 3 各分类模型在 Cifar-100 上的分类精确度比较

Table 3 Classification accuracy comparison of angle classification networks on Cifar-100 datasets

(单位:%)		
层类型	普通全连接层	角度全连接层
普通卷积层	76.25	75.79
SphereConv	77.99	78.04
角度卷积层	78.28	78.22

经过实验我们发现,使用角度特征的角度分类网络相比使用普通特征的 ResNet-32,分类准确率要高 2%,这充分论证了在分类模型中角度特征对于分类模型的重要意义。更加充分挖掘角度特征,或许是提高分类模型准确率的一个重要方向。

4.4 角度函数的选择

最后,基于 Cifar-100,我们研究了在角度卷积和角度全连接层中采用不同角度函数时的分类效果(角度卷积层和角度全连接层的角度函数保持一致),实验结果如表 4 所列。从表 4 中可以很清楚地看到,以 Log 函数作为角度函数得到的分类模型的准确率相比其他的角度函数都高,证明了 Log 函数作为角度函数时对小角度有更好的区分度,使得经过分类模型输出的特征点更趋于直线排列,从而有效地提高了分类准确率。

表 4 不同角度函数下角度分类网络在 Cifar-100 上的分类准确率

Table 4 Classification accuracy of angle classification networks on Cifar-100 under different angle functions

角度函数	分类准确率/%
$\cos(\theta)$	77.48
$1 - \frac{2}{\pi} \theta$	77.87
$\frac{1}{\theta}$	77.93
$-\log(\theta)$	78.22

结束语 经过实验验证,得到以下结论:1)传统的 Softmax Loss 本身可以看作一种隐式的角度度量方式,经 Softmax Loss 训练得到的特征点呈现角度分布;2)在分类网络中,使用角度卷积层和角度全连接层替换常规卷积层和常规全连接层得到的角度分类网络,输出的特征点更趋于直线排

列,分类准确率更高;3)在角度分类网络中,使用 Log 角度函数处理得到的角度特征物理含义明确,能有效地加强小角度之间的区分度,提高模型准确率。由于角度特征对于分类准确率的提高有着重要的意义,充分挖掘角度特征将会是开发新的高精度分类网络的一个重要指导方向。

参 考 文 献

- [1] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. *Communications of the ACM*, 2017, 60(6): 84-90.
- [2] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. *arXiv*: 1409. 1556, 2014.
- [3] SZEGEDY C, LIU W, JIA Y Q, et al. Going deeper with convolutions[C] // 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA, USA: IEEE, 2015: 1-9.
- [4] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C] // 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA: IEEE, 2014: 580-587.
- [5] GIRSHICK R. Fast R-CNN[C] // 2015 IEEE International Conference on Computer Vision (ICCV). Santiago, Chile: IEEE, 2015: 1440-1448.
- [6] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137-1149.
- [7] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C] // 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA, USA: IEEE, 2015: 3431-3440.
- [8] SIMONYAN K, ZISSERMAN A. Two-stream convolutional networks for action recognition in videos[C] // *Advances in Neural Information Processing Systems*. 2014: 568-576.
- [9] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, 2016: 770-778.
- [10] HUANG G, LIU Z, MAATEN L V D, et al. Densely connected convolutional networks[C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2017: 4700-4708.
- [11] IOFFE S, SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[J]. *arXiv*: 1502. 03167, 2015.
- [12] SALIMANS T, KINGMA D P. Weight normalization: A simple reparameterization to accelerate training of deep neural networks[C] // *Advances in Neural Information Processing Systems*. 2016: 901-909.
- [13] LEI BA J, KIROS J R, HINTON G E. Layer normalization[J]. *arXiv*: 1607. 06450, 2016.
- [14] SCHROFF F, KALENICHENKO D, PHILBIN J. FaceNet: a unified embedding for face recognition and clustering[C] // 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA, USA: IEEE, 2015: 815-823.
- [15] LIU W, WEN Y, YU Z, et al. Large-margin softmax loss for convolutional neural networks[C] // *ICML*. 2016: 7.
- [16] RANJAN R, CASTILLO C D, CHELLAPPA R. L2-constrained softmax loss for discriminative face verification[J]. *arXiv*: 1703. 09507, 2017.
- [17] LIU W Y, WEN Y D, YU Z D, et al. SphereFace: deep hypersphere embedding for face recognition[C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI: IEEE, 2017: 212-220.
- [18] WANG F, CHENG J, LIU W Y, et al. Additive margin softmax for face verification[J]. *IEEE Signal Processing Letters*, 2018, 25(7): 926-930.
- [19] WANG H, WANG Y T, ZHOU Z, et al. CosFace: large margin cosine loss for deep face recognition[C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT: IEEE, 2018: 5265-5274.
- [20] DENG J, GUO J, XUE N, et al. Arcface: Additive angular margin loss for deep face recognition[J]. *arXiv*: 1801. 07698, 2018.
- [21] HUANG G B, LEARNED-MILLER E. Labeled faces in the wild: Updates and new reporting procedures: Technical Report UM-CS-2014-003 [R]. Massachusetts Amherst, Amherst, MA, USA, 2014.
- [23] LIU W Y, LIU Z, YU Z D, et al. Decoupled networks[C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT: IEEE, 2018: 2771-2779.



WANG Li-hua, born in 1995, postgraduate. His main research interests include computer vision and deep learning.



DU Ming-hui, born in 1964, professor, Ph.D supervisor. His main research interests include signal processing and image processing.