doi:10.3772/j.issn.1002-0470.2022.06.003

# 具有旋转运动模糊不变性的卷积神经网络: RMBI-Net<sup>①</sup>

郭锐②\*\*\* 郝优\*\*\* 许溟\*\*\* 贾丽\*\*\*\* 李华③\*\*\*

(\*中国科学院计算技术研究所智能信息处理重点实验室 北京 100190)

(\*\* 中国科学院大学 北京 100049)

(\*\*\*\*92728部队 北京100036)

(\*\*\*\*\*91977部队 北京100036)

摘要 针对高速旋转的相机拍摄图像产生的旋转运动模糊带来的目标分类较为困难的问题,本文主要研究手工特征与卷积神经网络(CNN)的结合,在网络结构底层赋予卷积神经网络不变性,提升网络在分类任务中的准确率。本文基于 Gaussian-Hermite(GH)矩旋转运动模糊不变量(RMB\_GHMI),通过计算卷积神经网络隐藏层特征图上的 RMB\_GHMI 来实现将旋转运动模糊不变性引入到卷积神经网络中,使网络本身具有一定的旋转运动模糊不变性,并使网络可从受到严重噪声干扰的旋转运动模糊的图像中直接进行目标分类。实验结果表明,在旋转运动模糊后的 MNIST 数据集上,相对于经典卷积神经网络,本文方法可以将图像分类准确率提升 30% 左右;在旋转运动模糊后的 CIFAR-10 数据集上,图像分类准确率可以提升 4%~16%。

关键词 卷积神经网络(CNN);旋转运动模糊;目标分类;Gaussian-Hermite(GH)矩;不 变量

0 引言

图像的旋转运动模糊在高速旋转的导弹导引头 所拍摄的图像中很常见,快速、准确地识别模糊目标 具有重要的意义。图像的旋转运动模糊是在相机成 像的曝光时间内及相机与目标相对旋转运动导致 的。近年来,卷积神经网络(convolutional neural network,CNN)及其衍生的网络在计算机视觉领域应用 非常广泛,极大地促进了该领域的发展,CNN 在许 多任务中表现远超过了传统的手工特征采样。例如 LeNet<sup>[1]</sup>、VGGNet<sup>[2]</sup>和 ResNet<sup>[3]</sup>等网络。这些网络 在特定任务中实现了非常好的性能。研究发现,对 于没有遵从经典采样原理的 CNN 衍生网络(如 VG-GNet、ResNet 等),其网络的泛化能力不能得到保 障,它们对于平移、尺度、旋转并不具有不变性<sup>[4]</sup>。 针对特定的应用场景,理想的网络应具有相应的不 变性,这会大幅降低训练的复杂度并提升网络性能。 研究人员提出了很多方法来解决神经网络中不变性 的问题。

解决 CNN 不变性最常用的方法是数据增强,该 方法简单有效,已经成为神经网络处理数据的一般 方法,然而该方法的缺点也很明显。首先是消耗大 量的计算时间和网络容量来对样本进行学习,同时 还会增加网络过拟合的风险。其次,这种不变性是 从数据中习得的,而不是网络本身的固有属性,对于 每一个新的任务,都要重新进行学习。其他用于训 练神经网络具有不变性的方法,大致可以分为以下 3 类。

① 国家重点研发计划(2019YFF0301801, 2017YFB1002703),国家重点基础研究发展计划(2015CB554507)和国家自然科学基金(61379082)资助项目。

② 男,1988 年生,博士生;研究方向:图像模糊不变特征;E-mail: hquguorui@qq.com。

③ 通信作者, E-mail: lihua@ict.ac.cn。 (收稿日期:2021-03-25)

第1类是增强输入的方法。这是数据增强的一 个变体,它通过对每一层卷积操作的输入进行增强, 而不是对输入数据进行增强<sup>[5-6]</sup>。这类方法可以使 CNN 获得全局不变性,但它们在本质上与数据增强 的方法没有太大区别,并不能降低训练的复杂度。 第2类是通过增强滤波器的方法来使网络获得不变 性。它们通过对每一个卷积层的滤波器进行旋转或 其他增强,来使网络获得不变性<sup>[78]</sup>。这类方法都 受滤波器的形状制约,不能将滤波器进行任意的操 作,且计算复杂度与前一种方法基本相当。第3类 是通过修饰滤波器来使网络获得不变性<sup>[9]</sup>。这类 方法引入了经典的计算机视觉如谐波滤波器、傅立 叶变换等来使 CNN 获得旋转不变性,但是显著增加 了计算成本。

本文主要研究手工特征提取与 CNN 的结合方法。基于 Gaussian Hermite(GH)矩旋转运动模糊不变量(rotational motion blur Gaussian-Hermite moment invariants, RMB\_GHMI)<sup>[10]</sup>,计算 CNN 隐藏层特征 图上的 RMB\_GHMI,从而实现将旋转运动模糊不变性引入到 CNN 中的目的,使网络本身具有一定的旋转运动模糊不变性。实验结果表明,该方法可以大幅改进典型 CNN 对于旋转运动模糊图像的分类 及识别准确率。

1 相关工作

许多研究致力于手工特征提取与 CNN 的结合, 它们大致可分为以下 3 类。

第1类是利用 CNN 作为特征提取工具,再将 CNN 特征与手工设计特征相结合。文献[11]将 CNN 提取的特征与手工特征输入到支持向量机中 用于诊断骨质疏松症。文献[12]提出了一种 CNN 与手工特征融合的框架识别户外人脸表情。

第2类是将手工特征集成到神经网络中,使网络具有不变量的性质。文献[13]提出了空间变换网络(spatial transformer network,STN),可以将相关的区域变换到同一标准形状,以此来减少学习过程中的形状外观的变化,提高学习效率。文献[14]提出了一种方向响应网络(oriented response networks,

ORN),将离散傅立叶变换引入到网络中来丰富卷 积模板的形式,使网络能够对不同的方向进行响应。 针对 CNN 中间层特征维度高、含噪声较多的问题, 文献[15]提出了 CNN 特征降维的方法,首先利用 主成分分析对 CNN 特征进行降维,在数据层面和人 类感知层面证明了其有效性;然后将降维后的 CNN 特征作为区域特征向量,利用多水平超像素分割和 随机森林回归构建了一个融合手工特征及降维 CNN 特征的显著性检测模型。

第3类是将手工特征的设计方法直接引入到 CNN中。文献[16]提出了卷积通道特征,将 CNN 特征和随机森林等方法相结合。文献[17]提出了 局部二值卷积,用几个重构的卷积核替代了局部二 值卷积中的中心插值的方法。文献[18]提出了 Gabor 卷积网络,将 Gabor 滤波器应用到 CNN 中,增强 了 CNN 对于方向和尺度的鲁棒性。文献[19]将尺 度不变特征变换(scale-invariant feature transform, SIFT)特征和 CNN 特征结合到一起,在统一的框架 中使用视觉特征来进行高效的检索。

# 2 基于 GH 矩的旋转运动模糊不变量

根据旋转运动模糊退化模型<sup>[10]</sup>,该类图像的形 成是对原图像作一系列旋转变换后的结果的叠加均 值。本节首先给出旋转运动模糊图像 GH 矩的定 义,再将该 GH 矩代入旋转 GH 矩不变量中筛选出 低阶次具有稳定不变性的旋转变换和旋转运动模糊 的 GH 矩不变量,即 RMB\_GHMIs,构成一组特征向 量应用于图像检索、目标识别等任务。图 1 所示为 典型的清晰图像和该图像旋转运动模糊后的示例。



图1 原图像与旋转运动模糊后图像示例

### 2.1 模糊图像的 GH 矩

结合旋转运动模糊的退化模型和 GH 矩的定

义,模糊图像 
$$g(x, y)$$
 的 GH 矩如式(1)所示。  
 $GHM_{pq}^{(g)} = \iint g(x_{(g)}, y_{(g)}) \hat{H}_{p}(x_{(g)}; \sigma) \hat{H}_{q}(y_{(g)}; \sigma)$   
 $dx_{(g)} dy_{(g)}$   
 $= \sum_{k=0}^{p+q} \{H(p,q,k) \cdot GHM_{k, p+q-k}^{(f)}\}$  (1)  
其中.

$$H(p,q,k) = \sum_{i=0}^{k} \left[ S(p,q,k,i) \cdot \frac{1}{T} \int_{0}^{T} (\sin^{p+k-2i}\theta_{(i)} \cdot \cos^{q+2i-k}\theta_{(i)}) dt \right]$$
$$S(p,q,k,i) = \left\{ \begin{pmatrix} (-1)^{k-i} \cdot C_{p}^{i} \cdot C_{q}^{k-i} & i \leq m \coprod k - i \leq q \\ 0 & \ddagger \end{pmatrix} \right\}$$

式中,  $GHM_{pq}^{(g)}$  为模糊图像的(p,q)阶 GH 矩,  $g(x_{(x)}, y_{(y)})$  为模糊图像,  $\hat{H}_n$ 为 GH 多项式。  $GHM_{k,p+q-k}^{(f)}$ 为原图像的 GH 矩, p,q,k,i 为正整数组 合数,H(p,q,k)为原图像 GH 矩的各项系数。旋转 运动模糊 GH 矩不变量的构建过程就是消除各项系 数的过程。S(p,q,k,i)为控制函数,当各个参数一 旦违反 *i*≤*p*、*k* - *i*≤*q* 的约束时,此项不符合实际意 义,系数为0。详细推导过程见文献[10]。

离散情况下模糊图像的 GH 矩如式(2) 所示。

$$ghm_{pq}^{(g)} = \sum_{k=0}^{p+q} \{H(p,q,k) \cdot ghm_{k,p+q-k}^{(f)}\}$$
(2)

其中,

$$\begin{split} H(p,q,k) &= \\ \sum_{i=0}^{k} \left[ S(p,q,k,i) \cdot \frac{1}{T+1} \sum_{i=0}^{T} (\sin^{p+k-2i} \theta_{(i)} \cdot \cos^{q+2i-k} \theta_{(i)}) \right] \\ S(p,q,k,i) &= \\ \begin{cases} (-1)^{k-i} \cdot C_{p}^{i} \cdot C_{q}^{k-i} & i \leq p \perp k - i \leq q \\ 0 & & \ddagger \& \end{cases} \end{split}$$

式中,  $ghm_{pq}^{(g)}$  为模糊图像的 (p,q)阶 GH 矩,  $ghm_{k,p+q-k}^{(f)}$ 为原图像的 GH 矩, H(p,q,k)为原图像 GH矩的各项系数。

## 2.2 旋转运动模糊图像的 GH 矩不变量

式(1)给出了模糊图像 GH 矩与原图像 GH 矩 的关系式,构建旋转运动模糊矩不变量的关键就是 要寻找各项系数之间的关系。

基于上述算法,文献[10]得到了5个四阶以内 平均相对误差在5%以内的不变量,组成了具有相 — 578 —

当稳定性的旋转运动模糊特征问量,记作 RMB\_  
GHMI-5,如式(3)所示。  

$$RMB_GHMI(1) = ghm_{20} + ghm_{02}$$
  
 $RMB_GHMI(2) = ghm_{40} + 2ghm_{22} + ghm_{04}$   
 $RMB_GHMI(3) =$   
 $ghm_{01}ghm_{03} + ghm_{01}ghm_{21} + ghm_{10}ghm_{12} + ghm_{10}ghm_{30}$   
 $ghm_{10}^2 + ghm_{01}^2$   
 $RMB_GHMI(4) =$   
 $ghm_{10} + ghm_{01}^2 - ghm_{10}ghm_{03} - ghm_{10}ghm_{21}$   
 $ghm_{10}^2 + ghm_{01}^2$   
 $RMB_GHMI(5) =$   
 $(ghm_{30} + ghm_{12})^2 + (ghm_{21} + ghm_{03})^2$   
 $ghm_{10}^2 + ghm_{01}^2$   
 $(3)$ 

其中, ghm 为图像的旋转运动模糊 GH 矩。

经过实验测试和对比,该特征向量对旋转变换 和旋转运动模糊具有很好的不变性和区分性,即便 在高斯噪声、椒盐噪声、泊松噪声、乘性噪声相当严 重的情况下,仍能保持良好的识别性能,且比起同类 方法,图像检测的准确率有了显著提升。当模糊增 大至人眼都无法辨别的程度,图像内容发生了巨大 改变, RMB GHMI-5 依然保持了很好的性能<sup>[10]</sup>。 因此,在旋转运动模糊图像检索和模式识别领域,基 于手工特征的旋转运动模糊 GH 矩不变量具有非常 重要的应用价值。

#### 3 **RMBI-Net**

本节将介绍旋转运动模糊不变卷积神经网络 (rotational motion blur invariance convolutional neural network, RMBI-Net)的构造过程,详细说明 RMB GHMI-5 是如何与 CNN 相结合,使得 CNN 具有了旋 转运动模糊不变性。

如式(3)所示,所有的 RMB GHMI 都可以表 示几个 GH 矩的线性组合。矩的计算本质上可以是 一种卷积运算,因此,可以将不变量改写为卷积形式 并通过神经网络实现。接下来给出 RMBI-Net 的构 造和训练策略。

构造 RMBI-Net 的思路为计算 CNN 隐藏层特征 图上的 RMB GHMI, 通过构建 RMBI 层实现 RMBI 操作。RMBI 层包含一组函数映射,输出每个通道的特征映射的 RMB\_GHMI-5。RMBI 模块如图 2 所示。



图 2 RMBI 模块示意图

图 2 中, RMBI 层为所实现的不变量 RMB\_GH-MI-5 的计算层。该层的输入可以是任意通道个数 的特征图, 对于特征每一个通道, 都可以看作是一个 二维图像, 对其进行不变量 RMB\_GHMI-5 的计算, 该层的输出为 n 组 RMB\_GHMI-5 组成的特征向 量。

基于 RMBI 层,可以将其结合到任意经典的神 经网络结构中,将其与神经网络结合,提升网络对于 旋转模糊变换的鲁棒性。RMBI-Net 基本的网络结 构如图 3 所示,将 RMBI 层应用到每一层的特征图 上,并将输出与网络最终的特征层级联,可以得到最 终的网络输出。RMBI-Net 的训练策略是所有的参 数都在一个端到端阶段进行训练和更新,全连接层 返回的梯度将为卷积滤波器的更新参数。而对于 RMBI 层,只向前传播到它的后续层,并不进行误差 的方向传播。通过对比不同的训练策略,端到端的 策略会使整个训练过程更加顺利。



图 3 RMBI-Net 端到端网络结构与训练策略示意图

神经网络的优势在于其可以进行有效的特征提 取,产生大量的特征图,并通过学习的方法得到较高 的准确率,但不变性是神经网络所缺乏的。不变量 的优势在于通过手工设计推导,可以精确计算得到 图像的旋转运动模糊的不变特征,但是不变特征的 维度是该方法的一个限制。本文提出的 RMBI-Net 的方法,通过计算神经网络特征图的不变量,一方面 可以增加不变量的个数,另一方面可以将不变性引 入到神经网络中,提高网络对于旋转运动模糊的鲁 棒性。之后将进行实验,对该方法的性能进行详细 分析。

# 4 实验与分析

本实验的目的是评估 RMBI-Net 对分类结果的 影响,因此,没有使用其他网络优化方案以及调整训 练超参数来提高准确率。实验运行平台中央处理器 为 Intel Core i7 7700,显卡为 GeForce GTX1080Ti,深 度学习框架为 TensorFlow 2 on Windows 10。为了保 证实验的客观性、可比性和再现性,本文选择较为基 础且典型的 LeNet<sup>[11]</sup>、Z2CNN<sup>[20]</sup>和 NIN<sup>[21]</sup>等 3 种 CNN 进行改进,并实现改进前后的对比实验(如 LeNet 与 RMBI-LeNet)。改进前后的网络使用相同 的超参。改进前 3 种网络的结构如图 4所示。



LeNet使用随机梯度下降法(stochastic gradient descent,SGD)训练策略,主要用于手写体字符的识别和分类,部分超参为(base\_hr:0.01,momentum: 0.9, weight\_decay:0.0005, hr\_policy:"inv", gam-- 579 -- ma: 0.0001, power: 0.75, max \_ iter: 10000)  $_{\circ}$ 

Z2CNN 使用 Adam 来求解,由 7 个卷积层(每 层有 20 个通道)、激活层(ReLU)、dropout 层以及归 一层组成。部分超参为(base\_hr:0.01, momentum: 0.9, momentum2: 0.999, delta:0.0001, weight\_decay:0.0001, hr\_policy:"step", gamma:0.1, stepsize: 15 000, max\_iter: 20 000)。

NIN 由一个全局池化层和 3 个 MLP 层组成。 部分超参为(base \_ lr:0.025, momentum:0.9, lr \_ policy:"step", gamma:0.05, stepsize: 10000, max \_ iter: 13000)。

### 4.1 基于 MNIST 的验证实验

MNIST<sup>[1]</sup>数据集包含 70 000 张从 0~9 的 28 × 28 张手写数字图像,其中包含 60 000 个训练样本和 10 000 个测试样本。基于 MNIST,本文构造了与其 对应的旋转运动模糊后的数据集 MNIST-rmb。数据 构造方法是对每一张图片进行 10°~90°随机角度 旋转运动模糊后,添加 sigma = 0.01 高斯噪声。变

换后的数据集记为 MNIST-rmb。原数据集 MNIST 与变换后的数据集 MNIST-rmb 示例如图 5 所示。





(a) MNIST
 (b) MNIST-rm
 图 5 MNIST 示例与 MNIST-rmb 示例对比

本实验分为2种训练策略,第1种为利用 MNIST-rmb训练集对改造前后的网络进行训练,然 后在 MNIST-rmb 的测试集上进行测试,上述3种常 规的 CNN 和改造后的 RMBI-Net 在改造前后的的混 淆矩阵对比如表1~3所示。可以看出,经过改造的 CNN 具有了很强的旋转运动模糊不变性,在控制实 验中,RMBI-Net 的准确率相比于原网络,提升了近 30%。充分说明了本文所提出改造方法的有效性。

924

准确率:

94.58%

92.21%

	由「	MNIST-rm	b 训练的	LeNet 在 I	MNIST-rm	b 测试集_	上的混淆知	巨阵		准确率
679	0	16	0	11	8	64	0	26	5	83.93%
11	1079	25	43	26	36	127	37	83	48	71.22%
8	10	607	79	6	10	7	21	35	0	77.52%
11	5	198	759	0	44	0	69	30	4	67.71%
10	0	15	2	600	67	12	77	13	137	64.31%
4	3	18	47	5	474	2	7	45	16	76.33%
214	4	34	8	147	43	701	6	45	76	54.85%
2	1	40	46	10	19	0	666	44	21	78.45%
22	33	76	25	36	129	15	81	606	103	53.82%
19	0	3	1	141	62	29	64	47	599	62.07%
									准确率:	66.70%
	由	MNIST-rm	ıb 训练的	RMBI-Lel	Net 在 MN	IST-rmb _	上的混淆知	巨阵		准确率
909	0	4	1	4	1	9	2	4	8	96.50%
41	1126	24	30	43	50	56	38	58	60	73.79%
10	6	951	7	6	5	1	15	2	2	94.63%
0	0	3	915	0	5	1	4	2	3	98.07%
6	0	3	3	891	2	7	13	2	7	95.40%
1	0	33	38	0	815	1	18	5	0	89.46%
13	0	3	1	12	6	882	3	1	2	95.56%
0	0	3	4	1	1	0	914	3	2	98.49%
0	3	6	5	2	6	1	3	894	1	97.07%

表1 LeNet 和 RMBI-LeNet 在 MNIST-rmb 上的混淆矩阵对比

0

0

2

6

23

1

0

18

3

郭 锐等	:具有旋转运动模糊不变性的卷积神经网络:	RMBI-Net
------	----------------------	----------

-										
	F	h MNIST	-rmb 训练	的 Z2CNN	在 MNIS	Γ-rmb 上自	句混淆矩阵	年		准确率
722	0	6	2	7	2	61	0	17	7	87.62%
8	1078	22	24	21	23	116	27	68	35	75.81%
21	20	797	201	11	30	17	42	99	1	64.33%
8	7	77	629	2	34	3	32	19	3	77.27%
6	0	16	1	551	46	32	32	9	35	75.69%
12	3	12	64	11	511	8	6	45	14	74.49%
131	3	18	3	45	18	602	0	16	16	70.66%
3	3	29	62	12	21	0	707	34	15	79.80%
22	20	47	20	18	107	17	41	524	36	61.50%
47	1	8	4	304	100	102	141	143	847	49.91%
									准确率:	69.68%
	由 N	INIST-rm	b 训练的 ]	RMBI-Z20	CNN 在 MI	NIST-rmb	上的混淆	矩阵		准确率
909	0	5	0	1	1	8	1	3	10	96.91%
34	1129	22	26	36	42	54	31	52	56	76.18%
20	4	956	9	19	9	7	8	4	7	91.66%
0	1	2	918	0	3	1	1	2	2	98.71%
6	0	3	2	901	4	4	15	2	7	95.44%
1	0	35	39	0	819	1	15	7	3	89.02%
10	0	1	0	10	4	883	3	0	2	96.71%
0	0	3	5	1	1	0	941	4	3	98.23%
0	1	5	9	2	8	0	7	898	2	96.35%
0	0	0	2	12	1	0	6	2	917	97.55%
									准确率:	92.71%

表 2 Z2CNN 和 RMBI-Z2CNN 在 MNIST-rmb 上混淆矩阵对比

第2种为仅利用 MNIST 数据集中的训练集对 改造前后的网络进行训练,在 MNIST-rmb 的测试集 上进行测试,上述3种常规的 CNN 和改造后的 RM-BI-Net 在改造前后的混淆矩阵如表4~6 所示。从 表中可以观察到,即便在没有学习旋转运动模糊图像的情况下,RMBI-Net的准确率相比于原网络依然提升了近24%~34%,证明了 RMBI-Net 对于原网络的大幅改善成效。

表 3 NIN 和 RMBI-NIN 在 MNIST-rmb 上的混淆矩阵对比

由 MNIST-rmb 训练的 NIN 在 MNIST-rmb 上的混淆矩阵 准确率													
796	0	9	2	15	11	104	0	26	12	81.64%			
7	1072	8	13	21	20	98	18	63	30	79.41%			
13	12	760	113	12	22	16	35	81	1	71.36%			
4	17	153	765	3	64	3	44	27	4	70.57%			
7	0	9	2	614	36	27	49	9	50	76.46%			
9	6	12	35	20	543	7	13	58	23	74.79%			
105	7	12	3	73	19	659	3	20	46	69.59%			
1	4	32	59	15	17	2	764	60	31	77.56%			
17	17	36	18	24	99	9	38	567	64	63.78%			
21	0	1	0	185	61	33	64	63	748	63.61%			
									准确率.	72.88%	-		

										续表3
	由	MNIST-r	nb 训练的	i RMBI-NI	N在MNI	ST-rmb 上	的混淆矩	阵		准确率
910	0	1	0	1	2	8	1	4	3	97.85%
34	1129	22	19	35	43	52	26	45	52	77.49%
17	2	950	5	22	6	7	9	4	11	91.97%
0	1	0	924	0	4	0	0	2	1	99.14%
6	0	5	5	908	3	5	14	4	9	94.68%
2	0	38	46	0	832	2	22	17	5	86.31%
11	0	2	0	7	1	884	3	0	0	97.36%
0	1	12	5	1	1	0	948	3	4	97.23%
0	2	2	0	1	0	0	1	892	2	99.11%
0	0	0	6	7	0	0	4	3	922	97.88%
									准确率:	92.99%

表 4 LeNet 和 RMBI-LeNet 在 MNIST-rmb 上的混淆矩阵对比

由 MNIST 训练的 LeNet 在 MNIST-rmb 上的混淆矩阵 准确率													
679	0	16	0	11	8	64	0	26	5	83.93%			
11	1079	25	43	26	36	127	37	83	48	71.22%			
8	10	607	79	6	10	7	21	35	0	77.52%			
11	5	198	759	0	44	1	69	30	4	67.71%			
10	0	15	2	600	67	12	77	13	137	64.31%			
4	3	18	47	5	474	2	7	45	16	76.33%			
214	4	34	8	147	43	701	6	45	76	54.85%			
2	1	40	46	10	19	0	666	44	21	78.45%			
22	33	76	25	36	129	15	81	606	103	53.82%			
19	0	3	1	141	62	29	64	47	599	62.07%			
									准确率:	46.58%			

由 MNIST 训练的 RMBI-LeNet 在 MNIST-rmb 上的混淆矩阵											
909	0	4	1	4	1	9	2	4	8	96.50%	
41	1126	24	30	43	50	56	38	58	60	73.79%	
10	6	951	7	6	5	1	15	2	2	94.63%	
0	0	3	915	0	5	1	4	2	3	98.07%	
6	0	3	3	891	2	7	13	2	7	95.40%	
1	0	33	38	0	815	1	18	5	0	89.46%	
13	0	3	1	12	6	882	3	1	2	95.56%	
0	0	3	4	1	1	0	914	3	2	98.49%	
0	3	6	5	2	6	1	3	894	1	97.07%	
0	0	2	6	23	1	0	18	3	924	94.58%	
									准确率:	70.42%	

以上 2 种训练策略生成的改造前后的网络准确 率如表 7 和表 8 所示,可以看到 RMBI-Net 比相对 应的典型 CNN 分类准确率有了显著提升,验证了本 文提出的 RMBI-Net 方法的优越性。

# 4.2 基于 CIFAR-10 的验证实验

为了进一步验证 RMBI-Net 的适用性,本文选 择更加复杂且真实的数据集 CIFAR-10<sup>[22]</sup>进行同上 的数据集变换生成 CIFAR-10-rmb,并执行同上 2 种

— 582 —

		125 1	DZ CININ 17F	NIVIDI-Z		WIN151-1		'f) /C (+ / )	20	
		由 MNIS	5T 训练的	Z2CNN ₹	E MNIST-1	mb 上的泪	昆淆矩阵			准确率
435	0	6	2	52	19	44	9	2	18	435
442	1133	384	328	473	486	524	433	539	552	442
51	0	442	35	21	17	7	10	2	13	51
0	0	0	399	0	0	0	0	0	1	0
0	0	1	0	387	0	1	5	0	2	0
4	2	189	235	1	368	2	133	34	5	4
23	0	3	0	11	0	379	2	2	6	23
25	0	4	10	34	2	1	436	7	22	25
0	0	3	1	0	0	0	0	387	0	0
0	0	0	0	3	0	0	0	1	390	0
									. 0 . 4. ) .	

表 5 Z2CNN 和 RMBI-Z2CNN 在 MNIST-rmb 上混淆矩阵对比

准确率: 47.56%

由 MNIST 训练的 RMBI-Z2CNN 在 MNIST-rmb 上的混淆矩阵 准确率												
760	0	4	1	5	4	66	1	17	7	87.86%		
6	1089	14	27	22	20	96	24	67	29	78.12%		
13	8	713	85	7	14	8	19	31	1	79.31%		
3	1	153	772	0	41	1	37	15	3	75.24%		
16	1	17	38	9	546	6	6	42	18	78.79%		
122	7	18	3	60	20	673	2	23	24	70.69%		
1	2	30	53	4	11	1	727	26	13	83.76%		
28	25	71	28	17	117	10	51	649	53	61.87%		
22	0	0	1	192	75	53	99	92	811	60.30%		
									准确率:	74.06%		

表 6 NIN 和 RMBI-NIN 在 MNIST-rmb 上混淆矩阵对比

		由 MN	HST 训练的	的 NIN 在	MNIST-rm	b 上的混	淆矩阵			准确率
438	0	4	1	61	19	52	9	6	5	73.61%
435	1130	308	280	449	447	501	363	508	529	22.83%
63	2	476	55	21	30	11	3	8	9	70.21%
0	1	0	397	0	2	0	1	1	2	98.27%
21	0	27	29	435	31	14	117	36	67	55.98%
2	2	201	229	2	360	2	100	19	4	39.09%
20	0	2	0	7	0	377	0	0	3	92.18%
0	0	7	17	4	0	0	432	9	5	91.14%
0	0	7	2	0	2	1	0	386	0	96.74%
0	0	0	0	3	1	0	3	1	385	97.96%
									准确率:	48.16%
		由 MNIST	`训练的 F	MBI-NIN	在 MNIST	-rmb 上的	J混淆矩阵	:		准确率
787	0	1	1	1	2	15	0	2	4	96.80%
173	1115	83	44	147	147	157	107	147	179	48.50%
9	4	837	8	16	1	1	16	4	1	93.31%
0	0	4	803	0	3	0	1	3	1	98.53%

2 0 6 6 795 5 1 31 11 5 92.23%

										续表6
4	16	66	108	3	719	12	36	26	7	72.12%
1	0	0	0	2	2	761	1	1	0	99.09%
2	0	1	4	0	1	0	1	774	0	98.72%
0	0	0	0	7	0	1	1	1	802	98.77%
									准确率:	82.77%

表 7 由 MNIST-rmb 训练的 CNN 和 RMBI-Net 准确率

	LeNet	RMBI-LeNET	Z2CNN	RMBI-Z2CNN	NIN	RMBI-NIN
准确率	67.70%	92.21%	69.68%	92.71%	72.88%	92.99%

表 8 由 MNIST 训练的 CNN 和 RMBI-Net 准确率

	LeNet	RMBI-LeNET	Z2CNN	RMBI-Z2CNN	NIN	RMBI-NIN
准确率	46.58%	70.42%	47.56%	74.06%	48.16%	82.77%

网络训练策略的验证实验。相比于 MNIST 数据集, CIFAR-10 是 RGB 图像,而 MNIST 是灰度图像;CI-FAR-10 的图片尺寸为 32 × 32,比 MNIST 28 × 28 稍 大;MNIST 是手写字符,CIFAR-10 含有现实世界中 的真实物体,不仅噪声很大,而且物体的比例、特征 都不相同。这些特点都为 CNN 分类带来更大的挑战。两种训练策略生成的改造前后的网络准确率如表9 和表 10 所示,可以看到 RMBI-Net 比原网络的分类准确率依然有明显提升。在更复杂的场景中,进一步验证了 RMBI-Net 的优越性。

表9 由 CIFAR-10-rmb 训练的 CNN 和 RMBI-Net 准确率

	LeNet	RMBI-LeNET	Z2CNN	RMBI-Z2CNN	NIN	RMBI-NIN
准确率	52.83%	57.13%	60.07%	64.56%	62.17%	78.92%

表 10	由 CIFAR-10	训练的	CNN 和	<b>RMBI-Net</b>	准确率
------	------------	-----	-------	-----------------	-----

	LeNet	RMBI-LeNET	Z2CNN	RMBI-Z2CNN	NIN	RMBI-NIN
准确率	31.18%	49.96%	48.28%	59.63%	54.31%	73.76%

#### 4.3 实验结果分析

在传统仅使用手工特征分类场景中,特征向量 的维数对于准确率有着至关重要的影响。尽管 RMB\_GHMI-5对于旋转和旋转运动模糊理论上是 具有不变性的,但是对于类似于 CIFAR-10 这样更 为复杂的场景,更多维的特征向量显然会有更好的 性能。但是在本文的方法下,由于有了神经网络强 大的学习能力,低维特征向量的劣势会减弱。同时, 传统的网络也会因为手工特征的加入,具备了一定 的旋转运动模糊不变性,显著增强了网络在此退化 情况下的鲁棒性。另外,由于 CNN 都是多层结构, 每一次在特征图上的不变量运算所得到的特征映射 都相当于增加了手工特征的维数。同样,类似本文 的方法也可以很容易地扩展到其他手工特征提取方 法与 CNN 的结合,实现不同应用场景下的网络增 强。

# 5 结论

本文提出一种将旋转运动模糊不变量与神经网

络结合的 CNN 架构。通过在隐藏层计算 RMB \_ GHMI-5 的特征图,将 RMB \_ GHMI-5 具备而神经网 络缺失的旋转运动模糊不变性融入到 CNN 的网络 结构中,并在不同数据集、不同网络训练方法的分类 实验中验证了 RMBI-Net 可以大幅提升经典 CNN 对 于旋转运动模糊图像的分类性能。该方法的缺点是 会增加计算复杂度,但可以通过设计更高效的计算 不变量的算法解决,使得网络效率进一步提升。另 外,通过引入更高维的特征向量,也有望进一步提升 改进后网络的性能。

#### 参考文献

- [1] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradientbased learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11):2278-2324
- [2] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [EB/OL]. https://arXiv.org/pdf/1409.1556.pdf:arXiv, (2015-04-10), [2021-03-18]
- [ 3] HE K, ZHANG X, REB S, et al. Deep residual learning for image recognition[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 770-778
- [4] AZULAY A, WEISS Y. Why do deep convolutional networks generalize so poorly to small image transformations? [EB/OL]. https://arXiv.org/pdf/1805.12177. pdf:arXiv, (2019-12-31), [2021-03-18]
- [5] 王金甲,杨中玉. 基于特征融合进行活动识别的 DC-NN 方法[J]. 高技术通讯, 2016,26(4):374-380
- [6] HENRIQUES J A F, VEDALDI A. Warped convolutions: efficient invariance to spatial transformations [C] // Proceedings of the International Conference on Machine Learning, Sydney, Australia, 2017:1461-1469
- [7] 陈志旺,王昌蒙,王莹,等. ECO 跟踪算法中 CNN 分层 插值及加权策略改进[J]. 高技术通讯,2020,30 (6):570-578
- [8] CHIDESTER B, ZHOU T, DO M N, et al. Rotation equivariant and invariant neural networks for microscopy image analysis [J]. *Bioinformatics*, 2019, 35 (14): 530-537
- [9] CHUI C K, LIN S B, ZHOUD X. Deep neural networks for rotation-invariance approximation and learning [EB/

OL]. https://arXiv.org/pdf/1904.01814.pdf:arXiv, (2019-04-03),[2021-03-18]

- [10] 郭锐,贾丽,郝宏翔,等. 基于 Gaussian-Hermite 矩的旋转运动模糊个变量[J]. 中国图像图形学报,2021
   (8):1-16
- [11] SU R, LIU T, SUN C, et al. Fusing convolutional neural network features with hand-crafted features for osteoporosis diagnoses[J]. *Neurocomputing*, 2020, 385: 300-309
- [12] REDDY G V, SAVARNI C V R D, MUKHERJEE S. Facial expression recognition in the wild, by fusion of deep learnt and hand-crafted features [J]. Cognitive Systems Research, 2020, 62; 23-34
- [13] JADERBERG M, SIMONYAN K, ZISSERMAN A, et al. Spatial transformer networks [C] // Proceedings of the 28th International Conference on Neural Information Processing Systems, Montreal, Canada, 2015: 2017-2025
- [14] ZHOU Y, YE Q, QIU Q, et al. Oriented response networks [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 4961-4970
- [15] WORRALL D E, GARBIN S J, TURMUKHAMBETOV D, et al. Harmonic networks: deep translation and rotation equivariance[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 5028-5037
- [16] YANG B, YAN J, LEI Z, et al. Convolutional channel features [C] // Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 2015: 82-90
- [17] XU J F, NARESH BODDETI V, SAWIDES M. Local binary convolutional neural networks [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 19-28
- [18] LUAN S, CHEN C, ZHANG B, et al. Gabor convolutional networks [J]. IEEE Transactions on Image Processing, 2018, 27(9): 4357-4366
- [19] ZHOU W, LI H, SUN J, et al. Collaborative index embedding for image retrieval [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40 (5): 1154-1166
- [20]方正,曹铁勇,郑云飞,等.高效深度特征提取及其在显著性检测中的应用[J].计算机辅助设计与图形学学报,2019,31(2):144-151

[21] LIN M, CHEN Q, YAN S. Network in network [EB/ OL]. https://arXiv.org/pdf/1312.4400.pdf; arXiv, (2014-03-04), [2021-03-18]

[22] HINTON G E, SRIVASTAVA N, KRIZHEVSKY A, et al.

Improving neural networks by preventing co-adaptation of feature detectors [ EB/OL ]. https://arXiv.org/pdf/1207.0580.pdf;arXiv,(2012-06-03),[2021-03-18]

# **RMBI-Net**: convolutional neural networks with rotational motion blur invariants

GUO Rui\*\*\*\*, HAO You\*\*\*\*, XU Ming\*\*\*\*, JIA Li\*\*\*\*\*, LI Hua\*\*\*\*

(\*Key Laboratory of Intelligent Information Processing, Institute of Computing Technology,

Chinese Academy of Sciences, Beijing 100190)

(\*\* University of Chinese Academy of Sciences, Beijing 100049)

(\*\*\*\* The PLA 92728 Unit, Beijing 100036)

(\*\*\*\* The PLA 91977 Unit, Beijing 100036)

#### Abstract

Aiming at the problem of object classification caused by the rotational motion blur of the image captured by the high-speed rotating camera, the combination of handcraft features and convolutional neural network (CNN) is studied to introduce the invariance into CNN and improve the accuracy of networks in classification tasks. In this paper, based on rotational motion blur Gaussian-Hermit(GH) moments invariants (RMB \_ GHMI), calculating RMB \_ GHMI on CNN feature map is used to introduce the invariance of rotational motion blur into CNN, which makes the network have certain invariance to rotational motion blur, and enables the network to classify objects directly from the image of rotational transform superimposed with rotational motion blur interfered by noises. Experimental results show that the proposed method can improve the classification accuracy of typical CNN on MNIST dataset image after rotational motion blur transforming by 30%, and improve the classification accuracy of typical CNN on CIFAR-10 dataset image after rotation motion blur transforming by 4% - 16%.

Key words: convolutional neural network (CNN), rotational motion blur, object classification, Gaussian-Hermite (GH) moment, invariant