

基于孪生卷积神经网络的图像融合^①



杨雪, 郑婷婷, 戴阳

(长安大学 信息工程学院, 西安 710064)

通讯作者: 杨雪, E-mail: 592806227@qq.com

摘要: 传统的图像融合算法多有计算复杂程度高、不能有效提取图像纹理等不足, 为了弥补以上传统算法, 提出了一种基于孪生卷积神经网络 (Siamese Convolutional Neural Network, Siamese CNN) 的图像融合方法. 首先, 用孪生卷积神经网络生成一个权重图, 该权重图包含了来自两个待融合图像的全部像素信息. 然后, 用图像金字塔对像素以多尺度的方式进行融合, 并且采用了局部相似性策略自适应调整分解系数的融合模式. 最后, 和现存的几种图像融合的方法进行了对比. 实验证明, 该方法有较好的融合效果, 具有一定的可实用性.

关键词: 卷积神经网络; 图像融合; 孪生网络; 孪生卷积网络

引用格式: 杨雪, 郑婷婷, 戴阳. 基于孪生卷积神经网络的图像融合. 计算机系统应用, 2020, 29(5): 196-201. <http://www.c-s-a.org.cn/1003-3254/7375.html>

Image Fusion Based on Siamese Convolutional Neural Network

YANG Xue, ZHENG Ting-Ting, DAI Yang

(School of Information Engineering, Chang'an University, Xi'an 710064, China)

Abstract: Traditional image fusion algorithms have many shortcomings, such as high computational complexity and inability to effectively extract image texture. To compensate these shortcomings of above traditional algorithms, an image fusion method is proposed based on the Siamese Convolutional Neural Network (Siamese CNN). First, we use the Siamese CNN to generate a weight graph, which contains all pixel information from the two images to be fused. Then, the image pyramid is fused in a multi-scale way, and the local similarity strategy is adopted to adjust the decomposition coefficient adaptively. Finally, several existing image fusion methods are compared. Experimental results show that the proposed method has sound fusion effect and is practical to some extent.

Key words: Convolutional Neural Network (CNN); image fusion; Siamese network; Siamese CNN

1 引言

图像融合是将在同一场景下的不同聚焦的两幅或者多幅图像的信息进行冗余互补, 从而形成一张具有完整信息的图像. 图像融合在许多领域都具有广泛的应用, 例如在医学医疗方面和遥感图像的融合等方面. 现如今, 存在的图像融合方法有很多种, 其中比较常用的方法是基于分辨率分解的图像融合, 基于分辨率分解的方法又有多种, 例如基于图像金字塔分解^[1,2]的方

法、基于小波变换^[3]的方法、基于相位一致性和方向对比度的非子采样轮廓线变换的融合方法^[4]、基于稀疏表示的同时进行正交匹配跟踪的融合^[5]方法等等, 但是上述的方法中的一些方法对采集的图像有很大的要求, 这就意味着对采集图像的传感器有着很大的负担, 而且有些方法在计算上很复杂. 基于以上不足, 本文提出了一种基于孪生卷积神经网络的图像融合方法.

卷积神经网络 (Convolutional Neural Network,

^① 收稿时间: 2019-09-27; 修改时间: 2019-10-22; 采用时间: 2019-10-30; csa 在线出版时间: 2020-05-07

CNN)^[6-8]是人工神经网络和深度学习这两者的完美结合,其通过反向传播算法^[9]来训练其中的权值,进而实现我们所想要的结果.卷积神经网络有很多优点,例如,有比较好的容错性能、自适应性能、比较好的自学能力等等.由于卷积神经网络以上的优点,将原来传统的卷积神经网络改进为孪生卷积神经网络,孪生卷积神经网络在实质上就是一个双通道的卷积神经网络,并且为了计算简单,本文算法将其中的一个卷积层换成了残差网络层.

2 Siamese 网络

“Siamese”网络是 Bromley J、Guyon I 等在 1993 年在第一次提出来的^[10].又在 2005 年由 Chopra S、Hadsre R 等^[11]进行了补充,在其文章中用孪生网络进行了人脸相似性判别,将 Siamese 网络分成两部分,前半部分用于特征提取,将两张图片输入进网络中输出两个特征向量,后半部分用构造的特征向量进行距离度量,计算输入图像的相似度以判别人脸的相似性.后又由 Zagoruyko S 和 Komodakis N 在文献^[12]中以 Siamese 网络为原始基础进行了改进,首次将空间金字塔池化 (Spatial Pyramid Pooling, SPP) 和 Siamese 网络进行了结合^[13].

Siamese 网络可以理解成是一种衡量相似性的算法,当要分类的类别很多但是每一个类别所对应的样本数目很少的时候可以用于分类. Siamese 网络在近年来多应用于人脸识别以及相似度检测.本文所提的方法是基于 Siamese 网络进行特征的提取从而生成一个特征向量. Siamese 网络有两个分支,并且两个分支的权重和结构是一样的. Siamese 网络具体的结构如图 1 所示,将原图片 X_1 、 X_2 送入到系统当中,判断 X_1 和 X_2 是否是同一图片, W 是系统共享的学习参数, $G_W\{X_1\}$ 和 $G_W\{X_2\}$ 是生成低维空间中的 X_1 和 X_2 的两点映射,这样就可以通过计算 $G_W\{X_1\}$ 和 $G_W\{X_2\}$ 两者之间的距离来判断 X_1 和 X_2 是否是同一张图片了.

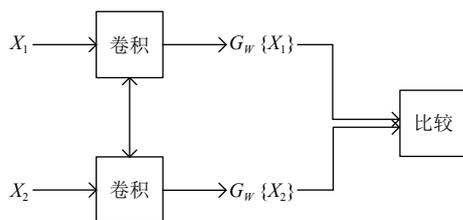


图 1 Siamese 原始网络结构图

3 本文所用网络

3.1 本文所用的孪生卷积神经网络

本文所采用的 Siamese 网络是文献^[12]中提及网络模型的改进版,本文网络用了两个完全相同的卷积神经网络,其中这两个网络是权值共享.

图 2 给出的是图像融合算法的前一部分特征提取的网络图,这是一个 Siamese CNN^[14]网络图,其中左右两边的两个部分在结构上是完全一样的,并且共享相同的参数,即权重和偏置.

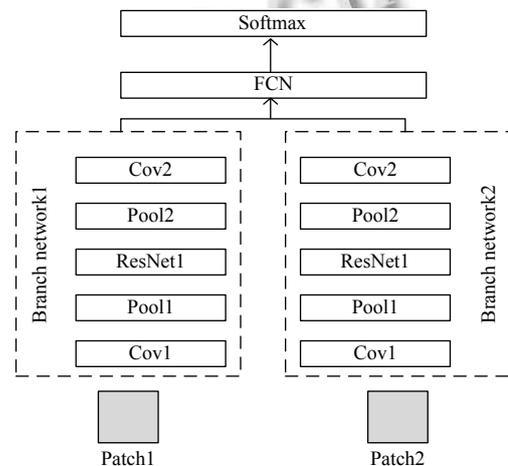


图 2 本文所使用的 Siamese 网络结构图

本文网络是由两个卷积层,两个池化层,一个残差块以及一个全连接层组成的.在本文中我们将原本孪生网络中的一个卷积层换成了残差网络^[15,16],所用的残差网络的结构如图 3 所示.输入的 patch 为 16×16 ,即卷积层和池化层的通道数都是 16(即卷积模板为 16 种),其卷积核的大小为 3×3 ,池化窗口尺寸为 2×2 ,卷积方式为零填充 (zero padding),卷积层的激活函数为 Relu,并且子 CNN 输出的是 256 维特征向量,每个池化层包括跨信道归一单元.首先将输入的 patch 的空间大小调整为 16×16 ,然后对网络进行训练以最大损失函数为优化目标(损失函数如式(1),采用随机梯度下降 (Stochastic Gradient Descent, SGD)^[17]的方法使其最大损失函数最小化;第二步将拼接完成的 512 个特征图直接连接在二维向量上面;最后,将连接生成的这个二维向量作为输入,输出给一个双向的 Softmax 层,其中 Softmax 可以用于分类,输出的是类别的概率分布,网络在 Softmax 层生成了两类的概率分布,这两个类别分别对应的是归一化之后的权重分配的结果.由于本文的网络是具有全连接的,所以输入和输出的数据

都是固定尺寸的,其模型配置如表1所示.

$$\|X_i - X_j\|_M^2 = (X_i - X_j)^T M (X_i - X_j) \quad (1)$$

其中, $i \neq j$, X_i 和 X_j 为两个子 CNN 的最后一层获得的两个 256 维特征向量, M 是一个正的半有限矩阵 ($M \in R^{256 \times 256}$).

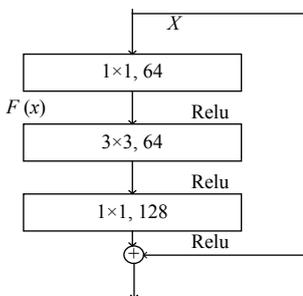


图3 残差网络结构图

表1 模型配置

模型	层	输入尺寸	卷积核/池化窗口尺寸	卷积核数量	步长	输出尺寸
Cov1	Cov1_1	16×16×1	3×3	64	1	16×16×64
	Cov1_2	16×16×64	3×3	128	1	16×16×128
Pool1	/	16×16×128	2×2	/	2	8×8×128
ResNet1	/	8×8×128	/	/	/	6×6×128
Pool2	/	6×6×128	2×2	/	2	3×3×128
Cov2	Cov2_1	3×3×128	3×3	256	1	3×3×256

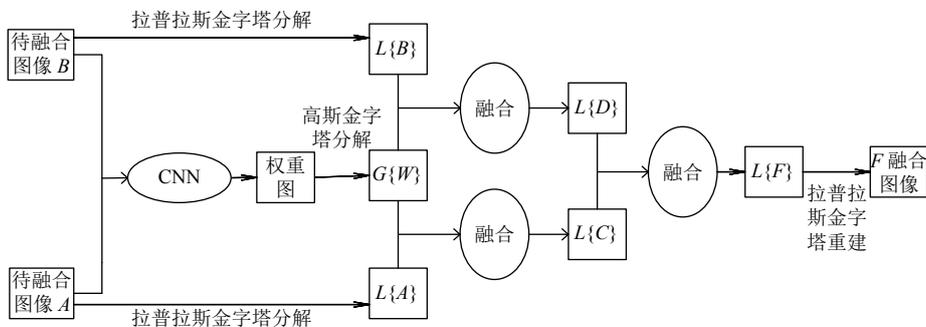


图4 融合过程图

对于每一个分解层,需要计算其 $L\{A\}^l$ 和 $L\{B\}^l$ 的局部能量:

$$\begin{cases} E_A^l(x,y) = \sum_m \sum_n L\{A\}^l(x+m,y+n)^2, \\ E_B^l(x,y) = \sum_m \sum_n L\{B\}^l(x+m,y+n)^2 \end{cases} \quad (2)$$

其中, $L\{A\}^l$ 和 $L\{B\}^l$ 分别表示了第 l 层拉普拉斯金字塔图像, $E_A^l(x,y)$ 、 $E_B^l(x,y)$ 分别表示了第 l 层,中心位置为 (x,y) 的 A 图和 B 图的局部区域能量,下同.

我们首先将全连通层转换为包含两个大小为 $3 \times 3 \times 512$ 的核的等效卷积层,转换后,网络可以将任意大小的原图像作为一个整体进行处理,生成一个密集预测图,其中每个预测(一个二维向量)包含对应位置的 patch 对的相对清晰信息.由于每个预测中只有两个维度,并且它们的和被标准化为1,因此输出可以简化为第一个(或第二个)patch 的权重.最后,为了得到与原图像大小相同的权值图,我们将该值赋值为 patch 位置内所有像素的权值,并对重叠的像素进行平均,最后得到最终的权重图.

3.2 融合过程

整个的图像融合过程可以分为两个部分,第一部分用孪生卷积神经网络进行图片信息的提取得到一个权重图,这一部分的详细描述在3.1章节;第二部分是利用金字塔分解后再融合,具体融合过程如图4所示,先将第一部分得到的权重图 W 经过高斯金字塔变换得到 $G\{W\}$,再将两张待融合图片 A 和 B 进行拉普拉斯变换分别得到 $L\{A\}$ 和 $L\{B\}$,再将得到的 $G\{W\}$ 和 $L\{A\}$ 、 $L\{B\}$ 分别进行融合,得到 $L\{C\}$ 和 $L\{D\}$,再将 $L\{C\}$ 和 $L\{D\}$ 进行融合得到 $L\{F\}$,最中将 $L\{F\}$ 进行拉普拉斯反变换得到最终的融合图像 F .

融合模式测定中的相似度测量可以表示为式(3):

$$S^l(x,y) = \frac{2 \sum_m \sum_n L\{A\}^l(x+m,y+n)L\{B\}^l(x+m,y+n)}{E_A^l(x,y) + E_B^l(x,y)} \quad (3)$$

其范围为 $[-1,1]$,若测量相似度为1,则说明两者具有高度相似性. t 表示了融合模式测定中的阈值,若 $S^l(x,y) \geq t$,则在加权平均融合时使用式(4):

$$\begin{cases} L\{C\}^l(x,y) = G\{W\}^l(x,y) \cdot L\{C\}^l(x,y) \\ L\{D\}^l(x,y) = G\{W\}^l(x,y) \cdot L\{B\}^l(x,y) \\ L\{F\}^l(x,y) = G\{W\}^l(x,y) \cdot L\{B\}^l(x,y) \end{cases} \quad (4)$$

如果 $S^l(x,y) < t$, 在融合时就选择比较 E_q 的局部能量, 如式 (5) 所示:

$$L\{F\}^l(x,y) = \begin{cases} L\{A\}^l(x,y), & \text{if } E_A^l(x,y) \geq E_B^l(x,y) \\ L\{B\}^l(x,y), & \text{if } E_A^l(x,y) < E_B^l(x,y) \end{cases} \quad (5)$$

最后, 对 $L\{F\}^l$ 进行拉普拉斯反变换得到最终融合图像 F .

4 实验结果

为了验证本文方法的有效性, 使用 3 组图像对网络模型的性能进行测试. 这 3 组图像分为 3 类, 第 1 类是医学图像组合 (如图 5 所示), 包括一张计算机断层扫描 (CT) 图像和一张核磁共振图像 (MRI); 第 2 类是风景图片 (如图 6 所示), 包括两张不同焦点的两张豹子的图片; 第 3 类是花草图像 (如图 7 所示), 包含了两张不同焦点的花草图像. 本文所有的实验均在 Lenovo CORE i5 Bth Gen 下的 Matlab 2016A 的版本下完成.

在实验中, 为更好测试本文方法的实用性, 选用两种现存图像融合方法和本文图像融合方法进行比较, 且比较了原始 Siamese CNN 网络结构的结果和改进后 Siamese CNN 网络的结果. 现存两种图像融合方法分别是基于导向滤波 (Guided Filtering, GF)^[18] 的图像融合方法、基于小波变换 (Wavelet Transform, WT)^[19] 的图像融合方法. 仅从医学图像的融合结果分析, 可明显看出来本文方法具有更高的对比性, 基于小波变换

方法的图像看起来存在很多噪声, 从视觉效果上来说不如本文方法. 3 种类型图片融合结果如图 8 至图 10 所示.

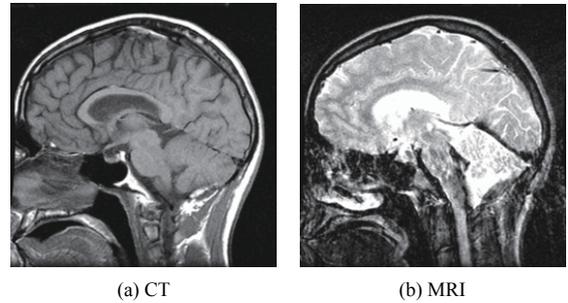


图 5 医学图像

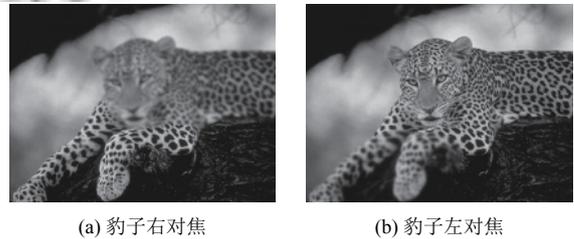


图 6 风景不同对焦图像

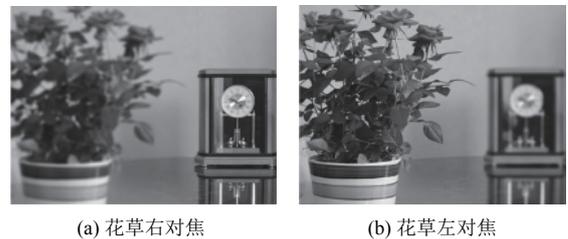


图 7 花草不同对焦图像

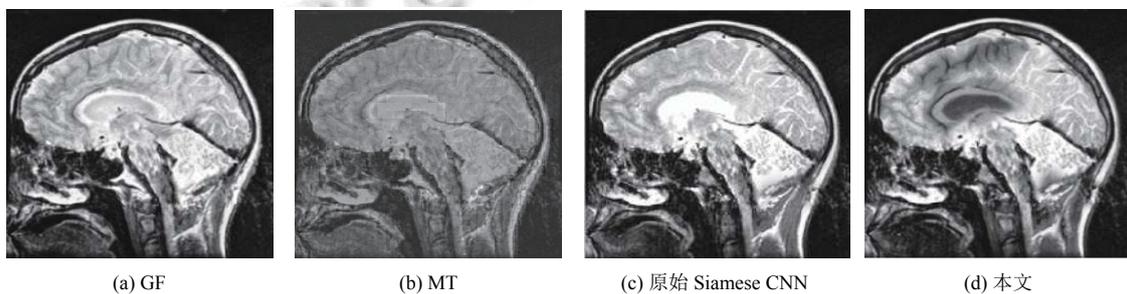


图 8 医学图像融合结果

由于本文实验图像是没有参考图像的, 为了能够更加客观的表示对比的方法和本论文的方法, 选取了几个客观的无参考图像情况下的图像融合评价标准. 第

一个是互信息 (MI), 互信息作为信息论中的一个很重要的概念, 可以用来度量两个变量之间的相关性程度, 在评价融合时可以表示融合之后的图像包含原始图像

的信息总量,这个值越大,就表示融合之后的图像包含的原始图像的信息越多,融合效果越好,在本文的实验中互信息的结果是融合图像分别和两个待融合图像的互信息之和;第二个是平均梯度 (AG), AG 描述了融合

之后图像的清晰度;第三个是边缘强度 (ES), ES 描述了融合之后图片的边缘信息, ES 越大就说明边缘信息越多. 最后给出了所有的方法在 CPU 上的运行时间以作参考. 对比结果如表 2 所示.

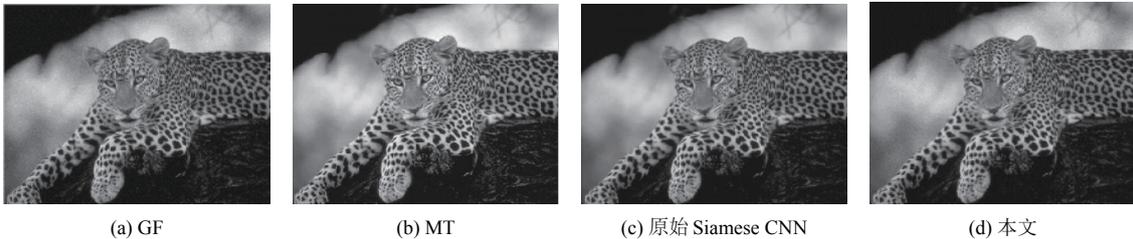


图 9 风景图像融合结果

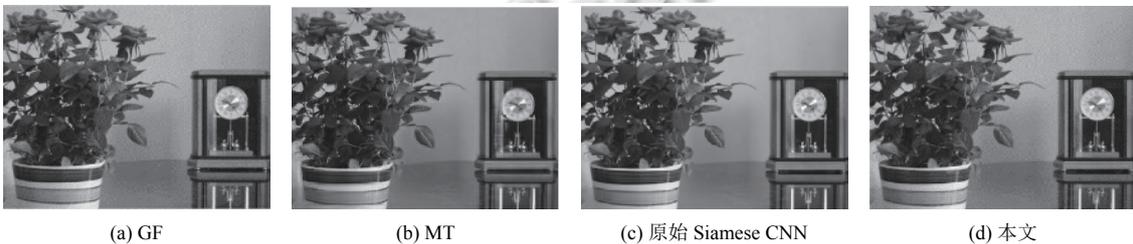


图 10 花草图像融合结果

表 2 对比说明

图像	方法	MI	AG	ES	T(s)
医学	GF	1.233	4.069	0.671	23.32
	WT	0.966	4.176	0.562	34.22
	原始 Siamese CNN	1.876	4.298	0.742	26.35
	本文	1.973	4.537	0.786	9.66
风景	GF	1.526	3.286	0.743	25.24
	WT	1.783	3.728	0.572	35.26
	原始 Siamese CNN	1.880	3.854	0.534	24.25
	本文	1.966	4.109	0.532	10.89
花草	GF	0.824	4.236	0.632	27.34
	WT	1.786	3.276	0.742	33.01
	原始 Siamese CNN	1.886	4.128	0.599	23.66
	本文	1.886	4.242	0.635	9.73

从表 2 中可见,本文所用方法融合之后的图像的互信息和平均梯度的数值都要大于用小波变换和导向滤波的数值,那就说明本文所提方法在融合之后的图像包含原图像的信息度和清晰度上都要优于小波变换和导向滤波的图像融合.但是在边缘信息的包含量上,在风景和花草这两类上面,本文所提的方法略逊于小波变换和导向滤波.但是单就视觉效果上来说,本文方法所得到的融合图像在视觉效果上要比其他两种办法好.与原始的网络结构相对比,在包含原图片的信息、

融合图片的清晰度、和融合图片的边缘信息上与改进之后的实验结果相差不多,但是在运行时间上却是大大的加快了,说明改进之后的方法在时效性上更好.综上所述,本文所提出的方法是具有一定的实用性的.

5 结论

本文提出了一种基于改进的卷积神经网络的像素级的图像融合方法,将原本孪生神经网络中的一个卷积层换成了残差块,比原本的算法更为简略,并且实验所用的时间更短,实现了像素级的图像融合.首先,用孪生卷积神经网络进行了图像的特征提取;然后,用图像金字塔对多尺度特征进行融合;最后,与现存的传统两种图像融合的方法进行了比较;侧面印证了本文方法的可实施性和可取之处.但是本文所提方法融合之后的图像的边缘信息包含量不如基于小波变换的图像融合,在这一点上有待改进.

参考文献

- 李增祥,漆志亮,贾楠,等.基于小波图像金字塔的工件目标检测与定位研究.现代电子技术,2019,42(15):33-37.
- 余美晨,孙玉秋,王超.基于拉普拉斯金字塔的图像融合算

- 法研究. 长江大学学报(自科版), 2016, 13(34): 21–26.
- 3 刘羽, 汪增福. 结合小波变换和自适应分块的多聚焦图像快速融合. 中国图象图形学报, 2013, 18(11): 1435–1444. [doi: [10.11834/jig.20131106](https://doi.org/10.11834/jig.20131106)]
 - 4 王瑞瑞, 马建文, 石伟, 等. 结合非子采样轮廓变换和形态收缩算子的多源遥感影像配准. 农业工程学报, 2012, 28(19): 110–118. [doi: [10.3969/j.issn.1002-6819.2012.19.015](https://doi.org/10.3969/j.issn.1002-6819.2012.19.015)]
 - 5 杨勇, 闫钧华, 井庆丰. 融合图像显著性与特征点匹配的形变目标跟踪. 中国图象图形学报, 2018, 23(3): 384–398. [doi: [10.11834/jig.170339](https://doi.org/10.11834/jig.170339)]
 - 6 Khan A, Sung JE, Kang JW. Multi-channel fusion convolutional neural network to classify syntactic anomaly from language-related ERP components. *Information Fusion*, 2019, 52: 53–61. [doi: [10.1016/j.inffus.2018.10.008](https://doi.org/10.1016/j.inffus.2018.10.008)]
 - 7 Liu Y, Chen X, Cheng J, *et al.* A medical image fusion method based on convolutional neural Networks. *Proceedings of 2017 20th International Conference on Information Fusion*. Xi'an, China. 2017. 1–7.
 - 8 Zhang Y, Liu Y, Sun P, *et al.* IFCNN: A general image fusion framework based on convolutional neural network. *Information Fusion*, 2020, 54: 99–118. [doi: [10.1016/j.inffus.2019.07.011](https://doi.org/10.1016/j.inffus.2019.07.011)]
 - 9 Sun W, Huang CC. A carbon price prediction model based on secondary decomposition algorithm and optimized back propagation neural network. *Journal of Cleaner Production*, 2020, 243: 118671. [doi: [10.1016/j.jclepro.2019.118671](https://doi.org/10.1016/j.jclepro.2019.118671)]
 - 10 Bromley J, Guyon I, LeCun Y, *et al.* Signature verification using a “Siamese” time delay neural network. *Proceedings of the 6th International Conference on Neural Information Processing Systems*. San Francisco, CA, USA, 1993: 737–744.
 - 11 Chopra S, Hadsell R, LeCun Y. Learning a similarity metric discriminatively, with application to face verification. *Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. San Diego, CA, USA, 2005: 539–546.
 - 12 Zagoruyko S, Komodakis N. Learning to compare image patches via convolutional neural networks. *Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition*. Boston, MA, USA, 2015: 4353–4361.
 - 13 He KM, Zhang XY, Ren SQ, *et al.* Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(9): 1904–1916. [doi: [10.1109/TPAMI.2015.2389824](https://doi.org/10.1109/TPAMI.2015.2389824)]
 - 14 Wang B, Wang L, Shuai B, *et al.* Joint learning of Siamese CNNs and temporally constrained metrics for tracklet association. arXiv: 1605.04502v2, 2016.
 - 15 李成, 汪小凯, 章振原, 等. 基于深度残差网络的有色金属破碎料视觉识别方法. *有色金属工程*, 2019, 9(8): 127–134. [doi: [10.3969/j.issn.2095-1744.2019.08.018](https://doi.org/10.3969/j.issn.2095-1744.2019.08.018)]
 - 16 Wu ST, Zhong SH, Liu Y. Deep residual learning for image steganalysis. *Multimedia Tools and Applications*, 2018, 77(9): 10437–10453. [doi: [10.1007/s11042-017-4440-4](https://doi.org/10.1007/s11042-017-4440-4)]
 - 17 Sreela SR, Idicula SM. Action recognition in still images using residual neural network features. *Procedia Computer Science*, 2018, 143: 563–569. [doi: [10.1016/j.procs.2018.10.432](https://doi.org/10.1016/j.procs.2018.10.432)]
 - 18 Wang Q, Zhang L, Zou WB, *et al.* Salient video object detection using a virtual border and guided filter. *Pattern Recognition*, 2020, 97: 106998. [doi: [10.1016/j.patcog.2019.106998](https://doi.org/10.1016/j.patcog.2019.106998)]
 - 19 王华兵, 邱春荣, 宫蓉蓉. 基于小波变换的图像融合算法比较研究. *长沙民政职业技术学院学报*, 2018, 25(3): 129–131.