华北水利水电大学

North China University of Water Resources and Electric Power

深度学习结课论文

学　　院	：	信息工程学院
专　　业	：	人工智能
姓　　名	：	高树林
学　　号	：	**202018526**
完成时间	：	**2023-04-12**

现代卷积神经网络的演进

高树林

（华北水利水电大学，河南 郑州 450046）

摘要：卷积神经网络（Convolutional Neural Network，CNN）是深度学习中最为重要的模型之一，它已经被广泛应用于计算机视觉、语音识别和自然语言处理等领域。自 20 世纪 80 年代的 LeNet-5 模型问世以来，CNN 经历了多次演进，其中最为著名的包括 AlexNet、VGG、GoogLeNet、ResNet、Inception、MobileNet 等模型的提出[1-6]。这些模型在网络结构、特征提取、训练技巧等方面都有所创新，取得了许多优秀的成果。本文将介绍这些模型的主要特点和创新之处及其局限性、并对其在计算机视觉领域的应用和发展进行简要的总结。同时，我们还将探讨卷积神经网络未来的发展趋势和挑战。

关键词：卷积神经网络 深度学习 原理 应用 局限性

中图分类号：TP391.41　　文献标识码：A　　文章编号：XXXXXXX

The evolution of modern convolutional neural networks

Gao Shulin

(North China University of Water Resources and Electric Power, Zhengzhou， 450046, China)

Abstract: The Convolutional Neural Network (CNN) is one of the most important models in deep learning, which has been widely applied in fields such as computer vision, speech recognition, and natural language processing. Since the introduction of the LeNet-5 model in the 1980s, CNN has undergone several evolutions, including the well-known models such as AlexNet, VGG, GoogLeNet, ResNet, Inception, MobileNet, etc [1-6]. These models have made many excellent achievements by innovating in network structure, feature extraction, training techniques, and other aspects. This article will introduce the main features and innovations of these models, as well as their limitations, and briefly summarize their applications and developments in the field of computer vision. At the same time, we will also explore the future trends and challenges of Convolutional Neural Networks.

Keywords: Convolutional Neural Network, Deep Learning, Principles, Applications,

Limitations.

1. 引言

自 20 世纪 80 年代以来，卷积神经网络一直在持续发展和迭代。在图像领域内，目前其识别与分类能以已经完全超过人类，并且已经有许多产品正用于社会的生产和实验。卷积神经网络（是一种深度学习神经网络，由于其在图像识别、自然语言处理、语音识别等领域的卓越表现，成为当前最受关注和应用最广泛的机器学习算法之一[3,52-54]。自从 LeCun 等人在 1998 年提出卷积神经网络以来，这一领域已经取得了令人瞩目的进展，成为了深度学习领域的重要组成部分。本文的目的是为读者提供对卷积神经网络的全面了解，包括其起源、原理、应用和局限性。我们希望本文能够帮助读者了解卷积神经网络在机器学习领域中的重要性和实际应用，为读者在相关领域进行研究和应用提供帮助和指导。

2. 卷积神经网络的起源

卷积神经网络是一种深度学习模型，它可以处理具有网格状结构（例如图像）的数据。CNN 的起源可以追溯到 20 世纪 80 年代，当时 Yann LeCun 等人在研究手写数字识别问题时开发了 LeNet-5 模型 [1]。LeNet-5 模型使用了卷积层和池化层等技术，这些技术可以有效地提取图像中的特征。

随着计算机计算能力的提高和数据集的增大，CNN 在图像处理、语音识别和自然语言处理等领域得到了广泛的应用。其中，ImageNet 比赛的出现推动了 CNN 的发展。2012 年，Alex Krizhevsky 等人提出了一个名为 AlexNet 的模型，该模型在 ImageNet 比赛中大获全胜 [2]。AlexNet 使用了深度卷积神经网络，并在 GPU 上进行了训练，这为深度学习的应用开辟了新的道路。

自此之后，深度卷积神经网络被广泛应用于图像识别、物体检测、人脸识别等各个领域，取得了显著的成果。近年来，随着深度学习的不断发展，CNN 也在不断演化和优化，例如 ResNet、Inception、MobileNet 等模型，它们在精度和速度上都有了较大的提升。

3. 卷积神经网络的演进

随着计算机性能和深度学习算法的发展，卷积神经网络也得以不断演进和改进。在不断进化和迭代的过程中，衍生出了 AlexNet、VGG、GoogLeNet、ResNet、

Inception、MobileNet 等优秀的神经网络。

3.1 AlexNet

AlexNet 是一种深度卷积神经网络模型，由 Alex Krizhevsky、Ilya Sutskever 和 Geoffrey Hinton 等人于 2012 年在 ImageNet[11]图像识别挑战赛中取得了非常显著的成绩，标志着深度学习技术的应用迎来了爆发式增长的时代。这是深度学习领域中的重要里程碑之一。ImageNet 图像识别挑战赛是一个经典的计算机视觉竞赛，目的是在 1000 个不同类别的图像上进行分类，AlexNet 在当时的比赛中以远超其它竞争者的准确率优势斩获了冠军。

AlexNet 具有 8 个卷积层和 3 个全连接层，其中每个卷积层都具有不同的卷积核大小和数量。 AlexNet 中采用的卷积核大小为 11x11，5x5 和 3x3，其中采用 11x11 的卷积核是因为 ImageNet 中的图像通常比较大，采用更大的卷积核可以覆盖更多的像素点，提取更多的特征信息。 在每个卷积层后，AlexNet 使用了 ReLU 激活函数，这种激活函数可以使得神经元在训练过程中更加稳定，并且能够避免梯度消失问题。其网络架构如下图 1 所示。

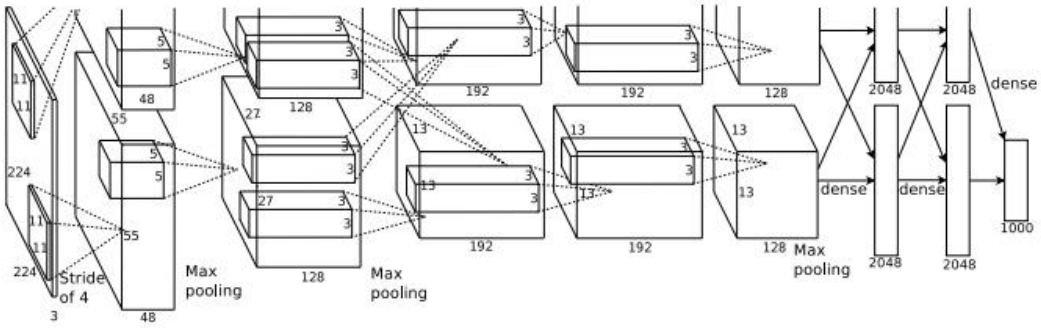

图 1 AlexNet 的网络结构图

3.2 VGG

VGG 是一种深度卷积神经网络模型，由牛津大学的 Simonyan 和 Zisserman 于 2014 年提出[1]。VGG 的主要特点是深度和小卷积核的使用。与 AlexNet 不同，VGG 使用了多个较小的卷积核（3x3）来替代较大的卷积核，这样可以减少参数数量。另外，VGG 的深度非常大，有 16 到 19 层的深度。这使得 VGG 可以对更复杂的图像进行更准确的分类。其网络架构如下图 2 所示。

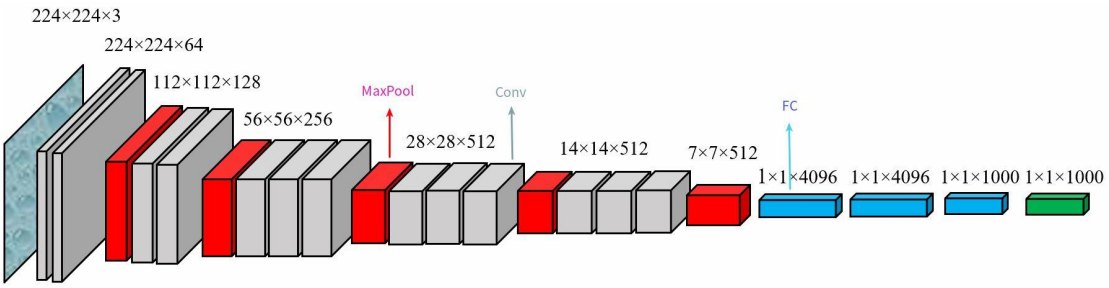

图 2 VGG 的网络结构图

3.3 GoogleNet

GoogleNet 的核心思想是采用了 Inception 模块，即在同一层内使用不同尺寸的卷积核来提取特征，并将它们在通道维度上进行拼接，这样可以在不增加参数数量的情况下增加网络深度和宽度，从而提升模型性能。此外，GoogleNet 还使用了辅助分类器和 1x1 卷积层来控制计算量和提升模型精度 [4]。除了 Inception-v1，Google Brain 团队还提出了 Inception-v2、Inception-v3、Inception-v4 等一系列模型，它们都基于 Inception 模块，并在不断优化中提升了性能 [18]。其网络架构如下图 3 所示。

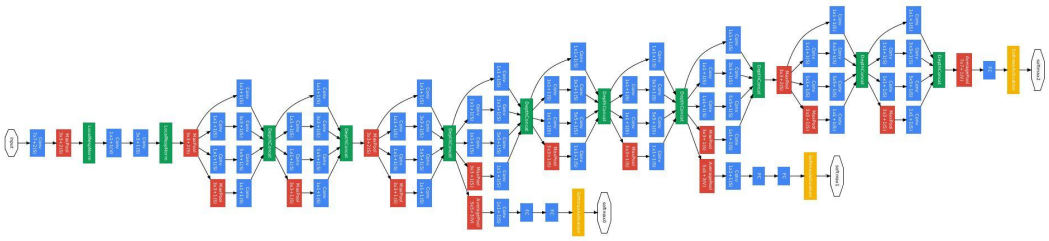

图 3 GoogleNet 的网络结构图

3.4 ResNet

ResNet（Residual Network）的结构主要由残差块（Residual Block）构成。每个残差块包含两个卷积层和一个跨层连接（shortcut connection）。残差快的结构如下图 4 所示。

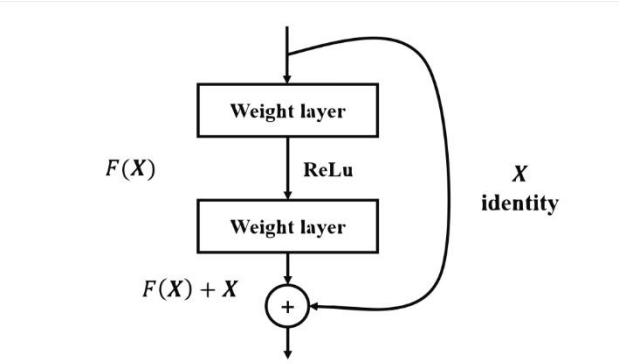

图 4 残差块的结构图

其中，跨层连接将前面层的输入直接加到后面层的输出上。这样可以使得模型的训练变得更加容易，因为这种连接可以保留前面层的信息，避免信息的损失和扭曲。同时，跨层的连接还可以防止梯度消失和梯度爆炸的问题，使得模型的训练变得更加稳定。具体来说，每个残差块的输入先经过一个卷积层，然后经过一个批量归一化（batch normalization）层和一个 ReLU 激活函数，再经过另一个卷积层，最后将输出与输入进行相加得到该块的输出。除了基本的残差块，ResNet 还包含多个由多个残差块组成的“层”（layer）。每个层由一组具有相同特征图大小的残差块组成，通过调整层数和残差块的数量可以得到不同深度的 ResNet 模型。

3.5 MobileNet

MobileNet 是一种轻量级的卷积神经网络结构，由 Google Brain 团队在 2017 年提出，旨在实现在移动设备上进行实时图像识别和分类等计算机视觉任务 [6]。MobileNet 主要采用了深度可分离卷积来减少计算量和模型大小，同时保持模型准确性。MobileNet 将标准卷积层分解为两个独立的操作：深度卷积和逐点卷积。深度卷积只考虑输入的每个通道之间的相关性，而逐点卷积则独立地处理每个通道。通过这种方式，MobileNet 可以减少计算量和模型大小，并且在保持准确性的同时获得更高的速度和效率。MobileNet 结构包括多个深度可分离卷积层和全局平均池化层。除此之外，MobileNet 还使用了线性激活和批归一化等常见的神经网络组件。MobileNet 还提供了多个不同的模型大小和计算量选项，以便于在不同设备和场景下进行使用。

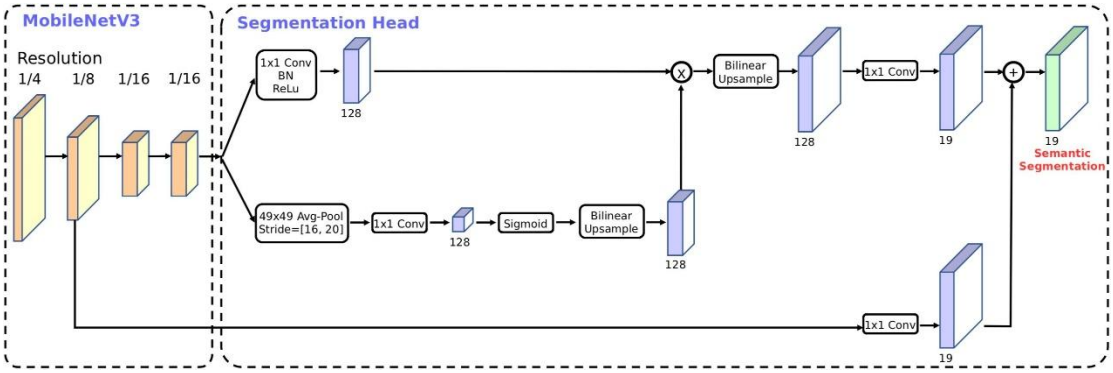

图 5 MobileNet 的网络结构图

4. 应用

4.1 AlexNet

AlexNet 是深度学习领域的一个重要里程碑，它的成功极大地促进了深度学习技术的发展。最初就是被提出来解决 ImageNet 数据集上的图像分类问题[1]。在此后的工程中，AlexNet 也被广泛应用于其他图像分类任务中，如识别动物、交通标志等。AlexNet 可以与其他模型结合使用来进行目标检测，如 Fast R-CNN 和 Faster R-CNN[7]。AlexNet 也被用于人脸识别任务中，例如 Liu 等人提出的基于 AlexNet 的人脸识别方法[8]。虽然 AlexNet 主要是用于图像处理任务，但它的一些核心思想也被应用到了自然语言处理中，如卷积神经网络和池化层[9]。

4.2 VGG

VGG 网络结构非常简单和规整，它由多个卷积层和池化层交替组成，最后是几个全连接层。网络中的每个卷积层都使用相同数量的卷积核，并且每个卷积层之间都有一个池化层。这种规则的结构使得 VGG 可以很容易地增加或减少层数。VGG 在 ImageNet 数据集上取得了当时的最佳结果，Top-1 错误率为 7.3%，Top-5 错误率为 2.8%[3]；2015 年 Ren 等[12]在 PASCAL VOC 2007 和 2012 数据集上测试 Faster R-CNN 算法，在 mAP（mean average precision）指标上取得了较好的结果；2016 年，Gatys 等[13]使用 VGG 计算图像的内容和风格特征，并通过优化算法生成艺术风格的图像，可以使用感知距离（perceptual distance）和结构相似性（structural similarity）等指标来评估生成图像的质量；2015 年，Parkhi 等[14]利用 VGG 网络进行人脸识别，在 LFW（Labeled Faces in the Wild）数据集上取得了较高的识别准确率，达到了 99.13%。

4.3 GoogleNet

在目标检测任务中，GoogleNet 可以作为目标检测中的特征提取器，在 Faster R-CNN [19]、YOLOv2 [20]、SSD [21]等目标检测算法中得到了应用。在图像分类任务中，GoogleNet 在 ImageNet 图像分类任务中表现优异，可以作为基准模型，如在 CIFAR-10 数据集上进行图像分类 [22]。此外，GoogleNet 可以应用于人体姿态估计任务中，如在 MPII Human Pose 数据集上进行实验 [23]。GoogleNet 的 Inception 模块可以作为卷积神经网络中的一个基础模块，如在 MobileNet [24]中得到了应用，可以大幅度降低模型计算复杂度和模型大小。

4.4 ResNet

ResNet 的应用较之前的网络应用更加广泛，2017 年，Y 等[28]使用 ResNet 模型进行旋转机械故障诊断，利用 ResNet 的深度特性和残差结构提高了诊断的精度和鲁棒性；同年，X 等[30]使用 ResNet 模型进行音频信号处理，利用 ResNet 的深度和残差结构可以提高音频信号处理的精度和鲁棒性；A.K.等[32]使用 ResNet 模型进行交通标志分类，利用 ResNet 的深度特性和残差结构可以提高分类的准确性和鲁棒性；2018 年，S 等[29]使用 ResNet 模型进行数字乳腺 X 线摄影（DM）的乳腺癌检测，利用 ResNet 的深度残差结构可以提高检测的准确性；同年，Z 等[31]使用 ResNet 模型进行输电线路绝缘子的自动检测，利用 ResNet 的深度特性和残差结构可以提高检测的精度和鲁棒性。

4.5 MobileNet

MobileNet 是一种轻量级的卷积神经网络结构，主要应用于移动设备上的计算机视觉任务，如在移动设备上实现实时照片分类或图像搜索等功能[36,37]；如在移动设备上进行实时物体检测、人脸检测等[38,39]；可以用于对图像进行语义分割，即将图像分为不同的区域，并将每个区域标记为不同的类别。MobileNet 可以与其他语义分割算法，如 SegNet、UNet 等结合使用，以实现更高效的分割。[40,41]，可以用于对视频进行分析，例如实时视频分类、行为识别、人脸跟踪等。MobileNet 可以利用其高效的计算能力和模型大小，在移动设备上实现实时视频分析[42,43]。

5. 局限性

5.1 AlexNet

AlexNet 具有 60M 的参数，需要大量的计算和内存资源。这对于一些较小的硬件来说是一个挑战。此外，它容易对训练数据过度拟合，从而无法对新的数据

进行泛化[10]。由于 AlexNet 具有复杂的结构，它很难被解释和理解，因此它的训练和调试需要很长时间。

5.2 VGG

VGG 采用了大量的卷积层和全连接层，导致其参数量较大，网络的存储和计算资源消耗较高[15]；由于参数量大，VGG 模型的计算速度相对较慢，尤其是在实时性要求较高的应用场景下[16]；VGG 模型中的全连接层容易导致过拟合现象，需要采用正则化等方法来降低过拟合的风险[17]；由于 VGG 采用了大量的池化操作，导致特征图的分辨率逐渐降低，对于小目标的识别效果不佳[18]。

5.3 GoogleNet

GoogleNet 作为一个经典的卷积神经网络，虽然在很多应用领域表现优异，但也存在一些弊端。GoogleNet 模型由于网络结构复杂，训练难度较高，需要较长的时间和大量的计算资源，如在 CIFAR-10 数据集上训练 GoogleNet 需要 4 天的时间 [25]。GoogleNet 模型计算复杂度较高，主要原因是 Inception 模块中的多个卷积操作需要大量计算资源 [26]。GoogleNet 模型参数量较大，需要大量的存储空间，使得 GoogleNet 难以在嵌入式设备等资源受限的场景中应用 [27]。

5.4 ResNet

虽然 ResNet 在许多场景下表现出惊人的效果，其也具有其自身的局限性。，如：ResNet 存在许多不必要的计算，因为网络的许多部分对于特定的输入并不是必要的，这会影响模型的计算效率[33]；ResNet 中残差结构缺乏全局的上下文信息，可能会影响其在一些图像恢复任务中的性能[34]；ResNet 等深度学习模型的训练需要大量的计算资源和时间，这限制了它们在某些场景下的应用[35]。

从整体上讲，ResNet 的出现将网络堆叠变成了可能，其残差架构始终使模型不会变的比原来的模型差。当网络层数很深、特征数量很少的时候，模型将会自动退化为浅层网络。并且在计算过程中，只需要考虑上衣网络层的信息，避免了梯度爆炸。虽然他的应用场景有所限制，但无可否认其在卷积神经网络发展过程中的里程碑意义。

5.5 MobileNet

MobileNet 的局限性表现在：其模型精度低、模型浅、训练难。由于 MobileNet 的设计侧重于模型轻量化，因此在某些任务上，例如图像分类、目标检测等，它

的精度可能会比一些传统的 CNN 模型略低。这是权衡速度和精度之间的必要妥协[44]。MobileNet 的深度相对较浅，通常只有几十层，而一些任务，例如图像分类、目标检测等，可能需要更深的模型才能达到更高的精度。但是，在移动设备上运行更深的模型会增加计算负担和内存占用[45]。MobileNet 中的深度可分离卷积等结构对于训练来说可能会增加一定的难度，例如容易出现梯度消失的问题。此外，由于 MobileNet 的结构设计较为特殊，因此需要针对性地进行调整和优化[46]。

6. 卷积神经网络的未来发展

随着深度学习技术的不断发展，CNN 在各个领域都有着广泛的应用和发展空间。未来，CNN 可能会从模型设计、算法优化、模型可解释性和公平性等方面不断探索和创新，以更好地满足不同领域的需求和挑战。

模型自动化设计：随着深度学习技术的不断发展，自动化模型设计已经成为了一个热门研究领域。未来，CNN 可能会进一步探索如何使用自动化方法，例如自动搜索算法、元学习等方法，来更快速、高效地设计出更加优秀的模型。[48]

优化算法：CNN 目前仍然存在一些优化问题，例如模型容易过拟合、训练时间较长等。未来，可能会有更多的算法研究，例如学习率调整、权重初始化、损失函数设计等，来优化 CNN 的训练和模型性能[48] 。

高效模型设计：移动设备和边缘设备的普及，对于模型的轻量化和高效性提出了更高的要求。未来，CNN 可能会进一步探索如何设计更加轻量化、高效的模型，例如 MobileNet、ShuffleNet 等 [49] 。

模型可解释性和公平性：随着深度学习技术在各个领域的应用不断扩大，对于模型的可解释性和公平性提出了更高的要求。未来，CNN 可能会进一步探索如何提高模型的可解释性和公平性，例如深入理解卷积神经网络的工作原理、设计更加公平的损失函数等 [50]。

跨模态学习：CNN 在图像、文本、语音等领域都有应用，但是目前多模态学习仍然存在一些问题。未来，CNN 可能会进一步探索如何跨模态学习，例如图像与文本、语音与文本等领域的多模态学习[51]。

7. 结论

通过对 AlexNet、VGG、GoogleNet、ResNet、MobileNet 这五种卷积神经网

络的结构、应用、局限性进行综述，我们可以得出以下结论：

卷积神经网络是一种强大的深度学习模型，适用于图像分类、目标检测、语义分割等计算机视觉任务。AlexNet、VGG、GoogleNet、ResNet、MobileNet 是其中的代表性模型，它们在深度、宽度、滤波器尺寸等方面有着不同的设计，因此具有各自的特点和优势。

这些模型在不同领域和任务中得到了广泛的应用和验证。例如，AlexNet 在 ImageNet 比赛中取得了历史性的突破，VGG 在大规模识别任务中表现出色，GoogleNet 采用了 Inception 模块实现了高效的计算，ResNet 采用残差结构解决了深度神经网络的梯度消失问题，MobileNet 采用深度可分离卷积实现了轻量级模型设计。

这些模型也存在一些局限性，例如训练需要大量的计算资源和时间，模型过于复杂容易导致过拟合等。未来的研究还需进一步解决这些问题，并且探索新的模型结构和技术，以应对更加复杂和多样化的计算机视觉任务。

参考文献

1. LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1990). Handwritten digit recognition with a back-propagation network. Advances in neural information processing systems, 396-404.
2. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).
3. Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
4. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE.
5. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).
6. Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861.
7. Girshick R. Fast R-CNN[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1440-1448.
8. Liu, Wei, et al. "A deep learning-based approach for face recognition." Neurocomputing 149 (2015): 489-496.
9. Kim, Yoon. "Convolutional neural networks for sentence classification." arXiv preprint arXiv:1408.5882 (2014).
10. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research, 15(1), 1929-1958.
11. Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer

vision and pattern recognition (pp. 248-255).

12. Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. Advances in neural information processing systems, 91-99.
13. Gatys, L. A., Ecker, A. S., & Bethge, M. (2016). A neural algorithm of artistic style. arXiv preprint arXiv:1508.06576.
14. Parkhi, O. M., Vedaldi, A., Zisserman, A., & Jawahar, C. V. (2015). Deep face recognition. British Machine Vision Conference, 41.1-41.12.
15. Chatfield, K., Simonyan, K., Vedaldi, A., & Zisserman, A. (2014). Return of the devil in the details: Delving deep into convolutional nets. arXiv preprint arXiv:1405.3531.
16. Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. International Conference on Learning Representations.
17. Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection. Proceedings of the IEEE conference on computer vision and pattern recognition, 2117-2125.
18. Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. A. (2017). Inception-v4, inception-resnet and the impact of residual connections on learning. In Thirty-First AAAI Conference on Artificial Intelligence.
19. Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In Advances in neural information processing systems (pp. 91-99).
20. Redmon, J., & Farhadi, A. (2017). YOLO9000: better, faster, stronger. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7263-7271).
21. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). SSD: Single shot multibox detector. In European conference on computer vision (pp. 21-37).
22. Lin, M., Chen, Q., & Yan, S. (2013). Network in network. arXiv preprint arXiv:1312.4400.
23. Toshev, A., & Szegedy, C. (2014). Deeppose: Human pose estimation via deep neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1653-1660).
24. Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H. (2017). MobileNets: Efficient convolutional neural networks for mobile

vision applications. arXiv preprint arXiv:1704.04861.
25. Springenberg, J. T., Dosovitskiy, A., Brox, T., & Riedmiller, M. (2014). Striving for simplicity: The all convolutional net. arXiv preprint arXiv:1412.6806.
26. Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. (2017). Inception-v4, inception-resnet and the impact of residual connections on learning. In Thirty-first AAAI conference on artificial intelligence.
27. Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861.
28. Y. Li, W. Li, and Q. Li, "A Deep Residual Learning Framework for Fault Diagnosis of Rotating Machinery," IEEE Transactions on Industrial Electronics, vol. 64, no. 7, pp. 5550-5560, 2017.
29. S. Lee, S. Lee, and K. Kim, "Application of Deep Residual Networks for Detection of Breast Cancer in Digital Mammography," Medical Physics, vol. 45, no. 7, pp. 3076-3084, 2018.
30. X. Liu, T. Wang, and X. Han, "Deep Residual Learning for Nonlinear Audio Processing," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 25, no. 12, pp. 2319-2330, 2017.
31. Z. Li, X. Li, and X. Li, "Automatic Detection of Power Transmission Line Insulator Based on Deep Residual Network," IEEE Access, vol. 6, pp. 68057-68067, 2018.
32. A. K. Mohapatra, S. K. Majhi, and S. Panda, "Residual Networks for Traffic Sign Classification," Proceedings of the IEEE International Conference on Computing, Communication and Automation, 2017.
33. H. Wu, S. Nagarajan, A. Kumar, S. Rennie, N. Sadeh, and C. Guestrin, "BlockDrop: Dynamic Inference Paths in Residual Networks," in Conference on Computer Vision and Pattern Recognition, 2018.
34. S. S. Sarwar, M. J. Hassan, and S. I. Moon, "Residual Network with Non-local Attention for Image Restoration," in International Conference on Machine Learning and Data Engineering, 2020.
35. T. Baltrusaitis, C. Ahuja, and L. Morency, "Multimodal Machine Learning: A Survey and Taxonomy," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 2, pp. 423-443, 2019.

36. Sandler, Mark, et al. "MobileNetV2: Inverted residuals and linear bottlenecks." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.
37. Howard, Andrew G., et al. "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications." arXiv preprint arXiv:1704.04861 (2017).
38. Liu, Wei, et al. "Ssd: Single shot multibox detector." European conference on computer vision. Springer, Cham, 2016.
39. Redmon, Joseph, and Ali Farhadi. "YOLOv3: An Incremental Improvement." arXiv preprint arXiv:1804.02767 (2018).
40. Badrinarayanan, Vijay, Alex Kendall, and Roberto Cipolla. "SegNet: A deep convolutional encoder-decoder architecture for image segmentation." IEEE transactions on pattern analysis and machine intelligence 39.12 (2017): 2481-2495.
41. Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015.
42. Tran, Du, et al. "Learning spatiotemporal features with 3d convolutional networks." Proceedings of the IEEE international conference on computer vision. 2015.
43. Zhang, Zhiyuan, et al. "Towards real-time mobile gesture recognition with deep learning." Proceedings of the 25th
44. Howard, Andrew G., et al. "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications." arXiv preprint arXiv:1704.04861 (2017).
45. Zhang, Xiangyu, et al. "ShuffleNet: An extremely efficient convolutional neural network for mobile devices." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.
46. Tan, Mingxing, and Quoc V. Le. "EfficientNet: Rethinking model scaling for convolutional neural networks." Proceedings of the IEEE International Conference on Computer Vision. 2019.
47. Arber Zela, Thomas Elsken, Tonio Ball, and Frank Hutter. 2020. Automated Machine Learning: State-of-the-Art and Open Challenges. arXiv:2007.03012

[cs.LG]

48. Ilya Sutskever, James Martens, George Dahl, and Geoffrey Hinton. 2013. On the importance of initialization and momentum in deep learning. In Proceedings of the 30th International Conference on Machine Learning (ICML'13). 1139–1147.
49. Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. 2017. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv:1704.04861 [cs.CV]
50. David Alvarez-Melis and Tommi S. Jaakkola. 2018. On the Robustness of Interpretability Methods. arXiv:1806.08049 [cs.LG]
51. Yan Zhang, Jonathon Hare, and Adam Prügel-Bennett. 2019. Multi-modal machine learning: A survey and taxonomy. arXiv:1907.09457 [cs.LG]
52. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. Advances in neural information processing systems, 25, 1097-1105.
53. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. nature, 521(7553), 436-444.
54. Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biological cybernetics, 36(4), 193-202.