

基于卷积神经网络识别 β - β 衰变

张爽* 强雨†

School of Physics, Peking University

摘要

伴随计算资源和数据量的解放，深层神经网络取得巨大的成功，并因其强大的学习以及泛化能力而得到广泛的发展和应用。我们利用卷积神经网络来实现双 β 衰变实验中，太阳中微子本底和实验事件的分离。实验中，由于要采用千吨级别的液体闪烁体探测器，不可避免的会探测到来自太阳中 8B 衰变放出的中微子，并与实验信号在能量上产生堆叠。然而，两种信号具有不同的物理特征，双 β 衰变产生的光信号具有双轨迹特征，而本底事件的光信号只有单轨迹。获得的数据包括光信号的空间和时间信息，这些物理信息可以表征在数据集和图片上，从而利用神经网络对 $\beta\beta$ 衰变事件和太阳中微子进行区分。

Keywords: 卷积神经网络，双 β 衰变，二分类，预训练模型。

1 引言

数学理论的发展，使得神经网络结构从感知机到多层感知机，至深度神经网络，再至卷积神经网络和循环神经网络过渡。神经网络对现实的刻画能力越来越强，通过把信息分布储存到神经元中，具有很强的鲁棒性，可以通过对训练集的学习形成并利用强大的信息综合能力，表现出很好的泛化特性。而各种神经网络结构中，卷积神经网络应用最为广泛。启发于视觉神经科学，卷积神经网络通过卷积和池化学习图像中诸如：边缘、角点、纹理、形状等更复杂的特征，最终很好的形成图片的概念。另一方面，双贝塔衰变是目前物理研究的热，由于三代中微子是电中性以及质量远小于它们的轻子对，这些特性展示出中微子是马约拉纳粒子的可能性 [1]。而无中微子双 β 衰变则有助于揭开中微子是马约拉纳粒子还是狄拉克粒子之谜。双贝塔衰变使用Kiloton-Scale液闪探测，巨大的液体闪烁体可以探测到环境中弥漫的太阳中微子，对实验的信号造成干扰，通常需要对庞大的数据量进行数据分析来排除本底，这与神经网络对大数据集的需求不谋而合，并且卷积神经网络结构在图像识别和分类方面取得了巨大成功。

文章按以下结构展开：第二部分介绍数据集及数据集处理，第三部分简单的介绍了模型浅层卷积神经网络及VGG16预训练模型以及具体实现，训练流程和结果在第四部分给出进行描述，最后对结果做出讨论和展望。

*1701110103@pku.edu.cn

†1701110096@pku.edu.cn

最后，文章基于双 β 衰变实验数据特征，把它作为有监督的分类问题，通过卷积神经网络（CNNs）进行训练，来实现对探测器事例类别的预测，并取得了一定的结果。与此同时，尝试了在数据集上使用迁移学习，也获得了不错的结果和经验，并且当探测器给出的信号更精确、维度信息更多时，迁移学习将会更加明显的体现出其优势，并且我们的结果会取得提升。

2 数据集

数据通过CERN公布在Kaggle上的开放途径下载得到。数据集由14051775个探测器信号构成，每个信号对应球型探测器表面探测到一个光子，带有探测粒子的空间辐角和时间信息，共计8万次事例。

表 a: 事件0中的粒子信号及信息

EventID	Signal	Theta	Phi	Time
0	1	0.579059	6.127253	1.899265
0	2	2.181142	4.257784	0.513214
...
0	172	2.181142	4.343392	1.584476

事件中信号来自于两种途径：粒子在闪烁体中超光速引起的切伦科夫效应释放的切伦科夫光子；与闪烁体探测器中电子相互作用产生的各向同性的闪烁体光子。双 β 衰变和太阳中微子通过切伦科夫辐射产生光子的分布截然不同，理想情况下，双 β 衰变由于释放两个电子，两个电子释放方向相反的两个锥状光子束，而中微子释放的切伦科夫光则只有一个方向，如图 1中(a)图所示，给出了只有切伦科夫光子情况下的信号空间分布。

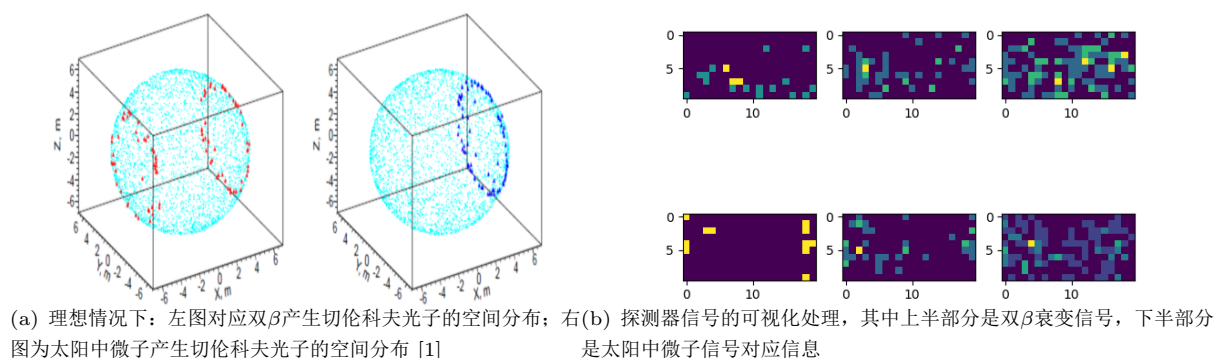


图 1: 粒子空间分布示意图及统计分布图

基于双 β 衰变与中微子事例的特征，我们以 θ - ϕ 为平面，对时间量做统计,按Bin区分时间(当作颜色通道处理),图 1中(b)图是按(10,20,3)统计所得，可以发现双 β 衰变信号具有显著的双轨迹特征的空间分布（沿 θ 分布较广），而中微子产生的切伦科夫辐射光信号则是单轨的（空间角的周期性，粒子聚集在 2π 附近）。

3 模型

3.1 浅层卷积网络

从20世纪60年代,受到猫脑皮层中的感知和方向选择神经元的独特的网络结构的启发,提出了卷积神经网络。至如今,卷积神经网络以其稀疏交互和局部权值共享的思想,在图像处理方面具有独特的优越性,被广泛的应用到众多领域,取得了优异的表现。

文中,基于双 β 衰变实验数据集的特征,我们构建了对实验数据进行分类的卷积网络结构模型 [2],选择如下卷积层特征:

1. 横向、纵向滑动步长均为1。
2. 3X3的滤波器
3. 保证卷积后长度工整,采用“SAME”方式填充边缘。

基于Tensorflow框架,构建了包含批归一化层 [3]和卷积层,池化层,全连接层的网络结构,非线性化除最后一层用sigmoid对输出做处理,其余层均用Relu做非线性化处理。选取交叉熵作为代价函数,为了防止过拟合,进行Dropout [4]处理及采用L2正则化,使用Adam [5]作为优化器;学习率初始设置为1e-3,按标准得初始化方式对参数进行初始化,对数据集按照8: 1: 1进行随机分割。mini-batch选为256。模型在Intel i7-7700K的处理器上运行。

3.2 基于VGG16的迁移学习

VGGNet由牛津大学的视觉几何组和Google DeepMind公司的研究员共同提出,是ILSVRC-2014中定位任务第一名和分类任务第二名 [6]。VGGNet统一使用卷积核大小为3X3,步长为1,填充方式为SAME和大小为2X2,步长为2的Max-Pool层,证明了使用小的卷积,增加网络深度可以有效提升模型的效果,而且VGGNet对其他数据集表现出很好的泛化能力,经常被用来提取图像特征。VGGNet探索了CNN的深度及其性能之间的关系,通过反复堆叠3x3的小型卷积核和2x2的最大池化层,VGGNet成功的构筑了16-19层深的CNN。而VGG16则包含16个卷积层和全连接层,以及1.38亿个参数,实际表现与VGG19相差不多。

由于硬件条件的限制,网络层数过深,计算量过大会导致堆栈溢出,很难从头训练一个较大的网络,无法将权重训练到最优。人的思维迁移能力让我们能够利用已有的知识,面临新的问题时,总是会找到与以往经验相似之处,很快的适应新的挑战,掌握新的方案,迁移学习便是受此启发,试图将模型从源任务上训练到的知识迁移到目标任务的应用上 [7] [9]。好在我们可以获得在更大型的数据集以及更好的设备基准下,重复训练给出的VGG16模型的权重参数。图像处理过程中,基本的特征都是一致的,VGG16模型对ImageNet数据集外的图片也表现出较好的泛化能力,我们可以采用训练好的VGG16来提取特征,应用到我们的问题上,来提高分类模型的表达能力。

对比ImageNet集,我们的数据集与原始数据相似度低,并且数据集只有8W张,我们尝试训练VGG16中特定层,冻结其它层的参数的想法。由于VGG16最初的几层网络捕获的是曲线、边缘这种广义的特征,显然是普适的,进行迁移的效果比较好 [8],可以冻结VGG16中前几个卷积层的参数,训练和微调后面层的权重和结构。

4 实践

4.1 卷积神经网络

通过histogramdd统计数据集里每个事例中粒子空间分布信息，按照不同Bin生成各种像素的图片(三维数组)，作为卷积层的输入。

表 b: 网络结构相同，不同大小输入的耗时

Pixel	Epoch	Train times/s
10 * 20 * 3	40	688
16 * 32 * 3	40	1793
32 * 64 * 3	40	7353

表中统计了3种输入，在包含5层卷积层和BN层，3层max pool层，3层全连接层的模型中，训练40个epoch的耗时。尽管没有严格的控制变量条件，我们仍记录了40个epoch时测试集的精度：分别达到81.9%，82.0%，82.3%，没有预期的随着像素的增高精度会有显著改善的趋势。并且当我们调整了网络结构，增加两个maxpool层减少参数数量，重新输入32 * 64 * 3，40个epoch的耗时仍旧高达6345s,并且精度上没有改善。

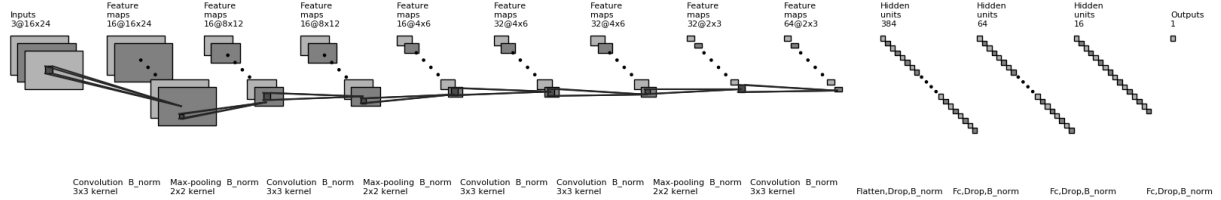
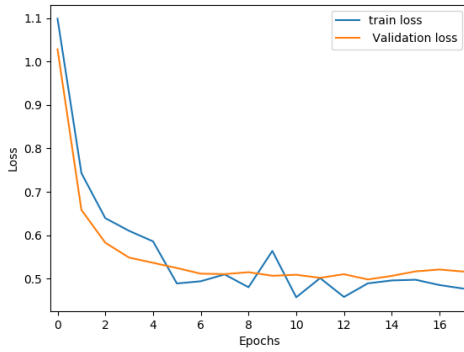


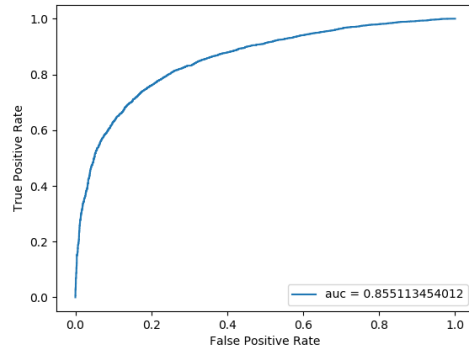
图 2: 模型结构示意图

考虑到时间因素和精度表现，最终选取按照16 * 24对 θ - ϕ 平面进行统计,并且通过TensorBoard可视化，发现各层权重参数随迭代逐渐变小在零附近，故选取了均值在0附近，方差为0.5的正态初始化条件，使用L2正则化只对权重参数进行惩罚，参数设置为0.005，学习率设置为1e-3，批归一化中尺度变换和偏移参数跟随卷积层和全连接层权重和偏置一起训练，初始化分别填1和0。

输入为16 * 24 * 3，使用图2所示的网络结构，我们得到的最好结果展示如下：



(a) 训练集和验证集的Loss与迭代次数关系图，横轴迭代整个训练集的次数，纵轴表示代价函数



(b) 测试集的ROC曲线，以及精度值

图 3: 浅层网络结构计算结果

验证集loss保持在0.49左右浮动，在测试集上，最终对双 β 衰变和太阳中微子本底事件的分类精度达到了85.51%，在未使用到事例的顶点重建 [10]（数据分析的重要组成部分,是物理分析的前提条件）信息的情况下，精度和ROC曲线已经好于球谐函数分析结果 [1]。并且直观统计了测试集所有事件的预测概率分布,如下图，可以看到由于loss仍旧比较大，有较大一部分事例预测正确但是概率分布在0.2-0.8之间，而且有较小的一部分预测错误，在尝试绘制统计图时也能发现有部分事例特征比较模糊，我们后续会尝试对数据集进行整理，以及寄希望于获得关于粒子更多的维度信息。

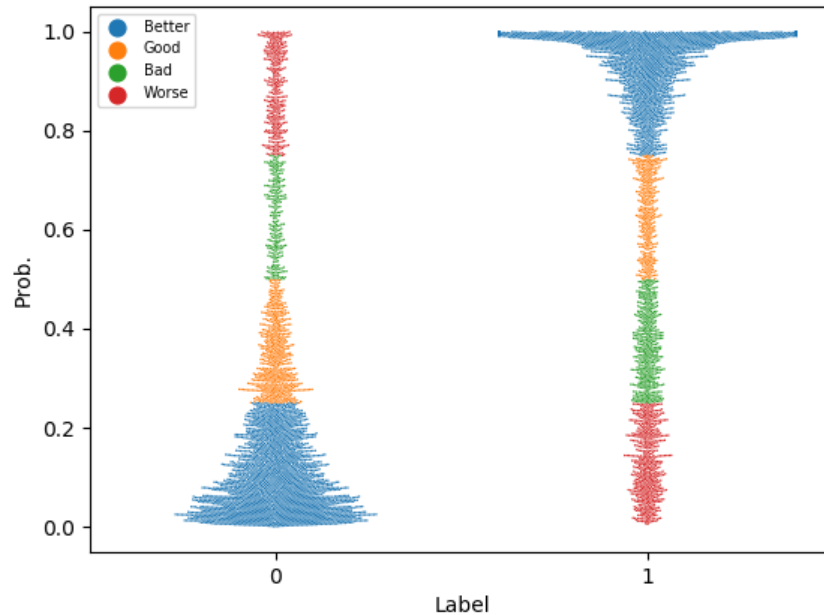


图 4: 横轴为二分类标签，纵轴为预测标签为1的概率，平铺的点对应真实事例数比，预测概率与真实结果的差值按阈值0.25分成四类，对应图中四种颜色，定量描述了预测准确程度。

4.2 VGG16预训练模型

使用预训练模型需要的输入为 $224*224*3$,有两种方法生成训练集数据:

1. 按Bin统计后产生相应大小图片,但会丢失掉大部分原始数组中信息。
2. 直接储存数组,但8万组数据要占用89.8G的储存空间。

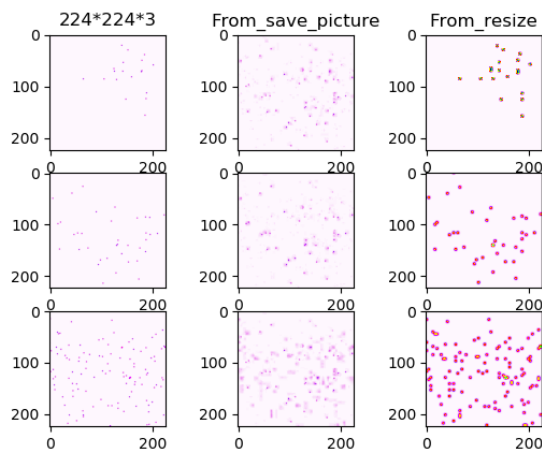


图 5: 每一行分别代表不同的channel, 第一列为 $224*224*3$ 输出的原始图片, 第二列为保存为图片后再读取的输出, 第三列对应储存输出 $56*56*3$ 信息后经resize的图片

通过在同一规格下的可视化比较, 第二种方案经过resize后和第一列原始数据更为相似。而且粒子数的统计随Bin增大更加趋于稀疏, 第三列还保留了小维度的优点。最终我们选择第二种方案, 批量读取时利用resize把数组伸张, 作为神经网络的输入。

预训练模型尽管提供了很大的方便, 但是要注意到与实际问题的出入, 由于我们的数据与原始数据相似度低, 且数据量较小, 尝试分别冻结VGG16前几个卷积层, 使用源任务训练出的权重作为特征提取器, 进行特征提取处理。由于输出pool2的储存量和计算量过于大, 首先尝试只冻结至pool3, pool4, pool5, 它们所需的储存空间分别为: 59.8GB, 29.92GB, 7.48GB。

受限向小调整了重新训练部分的网络结构, 并且pool3只随机抽取训练了4万个事例, 表c中只给出了卷积层部分的后置网络, 全连接层结构部分和图2中结构一致, 表c中给出了只重新训练全连接层和重新训练不同深度的卷积层的结果:

表 c: 调小后置网络训练预训练模型的具体训练信息和结果

预训练输出层	调整网络结构	batch	事例数	Loss	精度
pool3	Bn+conv-256+Bn+pool+conv-256+pool	128	4w	0.5	81.6%
pool4	pool+Bn+conv-512+Bn+conv-512+Bn+conv-512+pool	64	8w	0.5	79.7%
pool5	Bn+conv-512+pool+Bn	256	8w	0.5	73.8%

测试集的loss和精度均没有超前文中神经网络取得的结果，显然由于文中训练的数据集与ImageNet数据集间的差异，采用靠后卷积层参数及其学习到复杂的特征对应用到文中的数据集上没有帮助，故冻结至pool5层不重新训练卷积层取得了较差的结果。在pool3和pool4层的特征提取给出的结果上，虽然没有显著的改善，但并未打消我们使用迁移学习的想法，我们看到了一些改观的迹象，图 6中比较也可以看出，冻结全部卷积层和全部使用预训练模型提取的特征对我们的模型几乎没有起到分类的效果，而冻结至pool3和pool4层并加上后置重新训练的卷积层则表现出了一定的分类能力，并且pool3明显优于pool4的结果。

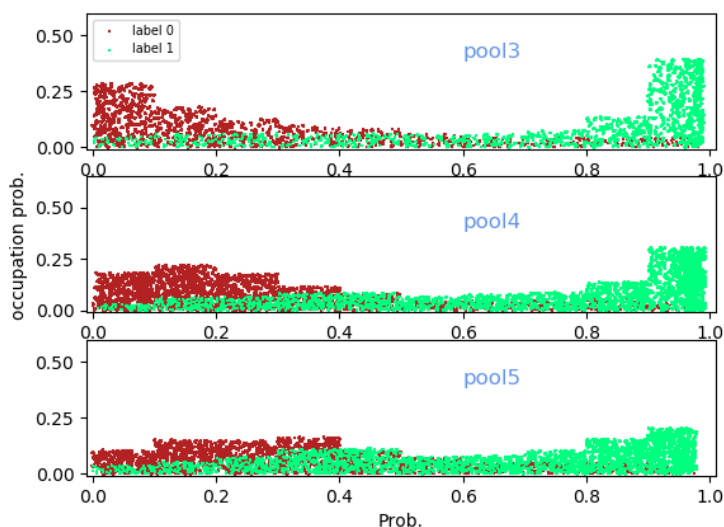


图 6: 预训练分类模型预测的测试集中事例概率的统计比较

在服务器上采用和VGG16一致的卷积层结构提取特征，并适当的冻结卷积层，由于模型很容易过拟合，于是通过early-stopping来避免过拟合问题，最终获得表中信息：

表 d: 采用VGG16卷积层结构，全连接层与前文一致，冻结至pool2,pool3,pool4,pool5的训练结果

预训练输出层	Epoch	事例数	Loss	精度
pool2	6	4w	0.5340	84.3%
pool3	7	4w	0.5396	83.8%
pool4	4	8w	0.5685	82.6%
pool5	6	8w	0.6632	70.8%

相比于之前的预训练神经网络，pool3、pool4对应的结果取得了显著的改善，pool2在4万的训练集达到了84%以上的精度。加了重新训练的卷积层提取特征后，pool2、pool3、pool4也取得了较好的结果。如图7，相比于上一小节模型，在预测的准确度上有了很大的进步。在预测的结果分布上，相

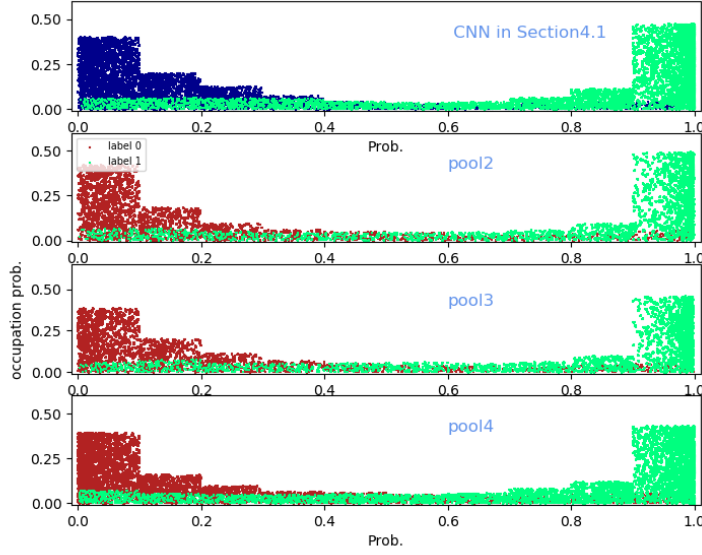


图 7: 预训练分类模型预测的测试集中事例概率的统计比较

比4.1节中模型，pool2在更小的数据集上取得了略优的结果，体现了迁移学习的可行性、有效性。并且在面对更精确、更多的探测器信号及其携带更多的维度信息时，迁移学习将更能体现出其优势所在。

5 结论与展望

应用卷积神经网络对双 β 衰变中事件信号进行分类处理，我们获得了85%的分类精度，相较于数据分析方法，神经网络模型不依赖顶点重建信息，获得了更好的结果。之后，将会对“时间”通道进行更精细的划分，并且尝试结合数据分析来帮助神经网络进行学习，例如加入顶点重建信息 [10]，以及采用更深的网络结构进行计算；在数据上，由于在产生图像特征时少量事例特征模糊，对这部分模糊数据进行处理或者筛除。并且训练所使用的数据集所含信息量较少，把模型应用于更精细的探测结果上，在获得更精准信号信息同时，以期获取更多诸如能量，速度分量，误差等信息，来提高神经网络的鉴别能力和精度。通过迁移学习，进一步了解了卷积网络通过逐级提取图像各个层次特征，来认知图像。通过参数迁移到文中的数据集上也取得了不错的结果，虽然源数据集和文中训练集差异较大，但通过VGG16预训练模型，冻结基础的特征提取层，获得了84.1%的精度。相信未来在更精准、事例内信息数量和维度更大的探测器信号数据集上，数据分析处理难度将会增大，但对于神经网络则是一大利好，而且迁移学习将会展示其明显的优势，及结果将获得显著的提升。

参考文献

- [1] ANDREY ELAGIN, HENRY FRISCH, BRIAN NARANJO, JONATHAN OUELLET, LINDLEY WINSLOW, TARITREE WONGJIRAD, *Separating Double-Beta Decay Events from Solar Neutrino Interactions in a Kiloton-Scale Liquid Scintillator Detector By Fast Timing*, <https://arxiv.org/abs/1609.09865>, (2016).
- [2] ALEX KRIZHEVSKY, ILYA SUTSKEVER, GEOFFREY E. HINTON, *ImageNet Classification with Deep Convolutional Neural Networks*, Advances in Neural Information Processing Systems **25**, <http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>, (2012).
- [3] IOFFE, SERGEY, CHRISTIAN SZEGEDY, *Batch normalization: Accelerating deep network training by reducing internal covariate shift*, <https://arxiv.org/abs/1502.03167>, (2015).
- [4] G. HINTON, A. KRIZHEVSKY, R. SALAKHUTDINOV, N. SRIVASTAVA, I. SUTSKEVER *Dropout: A Simple Way to Prevent Neural Networks from Overfitting*, Journal of Machine Learning Research **15**, (2014), pp. 1929–1958.
- [5] D. KINGMA, J. BA, *Adam: A Method for Stochastic Optimization*, 3rd International Conference on Learning Representations, <https://arxiv.org/abs/1412.6980>, (2015).
- [6] KAREN SIMONYAN, ANDREW ZISSERMAN, *Very Deep Convolutional Networks for Large-Scale Image Recognition*, <https://arxiv.org/abs/1409.1556>, (2014).
- [7] PAN, S.J., YANG, Q.: *A survey on transfer learning*, IEEE Trans. Knowl. Data Eng., 2010, **22**, (10), pp. 1345–1359
- [8] J. YOSINSKI, J. CLUNE, Y. BENGIO, AND H. LIPSON.: *How transferable are features in deep neural networks?*, In NIPS, 2014.
- [9] J TZENG E, HOFFMAN J, DARRELL T, ET AL., *Simultaneous deep transfer across domains and tasks*, Proceedings of the IEEE International Conference on Computer Vision, 2015: 4068-4076.
- [10] C. ABERLE, A. ELAGIN, H. J. FRISCH, M. WETSTEIN, L. WINSLOW, *Separating Double-Beta Decay Events from Solar Neutrino Interactions in a Kiloton-Scale Liquid Scintillator Detector By Fast Timing*, <https://arxiv.org/abs/1307.5813>, (2013).