



媒体与认知

医学图像分类

小组成员：（按照学号顺序排列）

李柘杙 2015011139

王博宇 2015011143

陈逸凡 2015011144

目录

一、摘要.....	2
二、问题介绍.....	2
三、Baseline 的复现.....	2
四、使用 Deep Mutual Learning 重新训练.....	3
1. 蒸馏学习介绍.....	3
2. Deep Mutual Learning 介绍.....	4
3. 训练结果.....	5
五、分类训练.....	6
1. 病理特征的介绍.....	6
2. 分类训练的结果.....	8
六、图像的预处理.....	9
七、可视化展示.....	10
八、实验总结.....	11
九、参考文献.....	11

一、摘要

医学图像的识别是一项十分重要的工作，其对病症的诊断与治疗都有十分重要的意义。我们这次实验是在 CheXNet 的工作的基础上，以提高对 ChestX-ray14 数据集上的近 11 万张胸片按照病症的不同进行分类的正确率为目标，尝试了几种不同的方法，最后实现和 CheXNet 几乎相同的正确率。我们尝试的方法有：利用 Deep Mutual Learning 进行网络的交叉训练，根据不同的疾病的特征不同对样本进行分类训练，根据胸片的特征对胸片进行预处理。最后我们完成了图形界面的开发和对病症进行可视化的标注。

Code and pre-trained models are available at:

<https://github.com/618clw/Medical-Images-Classification>

二、问题介绍

我们的任务是对 11 万张胸片完成多标签分类，类别有 14 种，分别代表 14 种不同的疾病。

三、Baseline 的复现

CheXNet 是一个使用了 Densenet-121 来实现对 ChestX-ray 数据集进行分类的分类器，其可以输出 14 种肺部疾病的有无。Baseline 除了使用了 DenseNet-121 复现了 CheXNet 的工作外，还使用了其它较为主流的网络结构来复现出了相似的结果。包括 ResNet-18, ResNet-50, DenseNet-169 等。考虑到网络的复杂程度以及训练时间的问题，我们用于训练的网络都是基于 ResNet-18, ResNet-50 和 DenseNet-121。值得一提的是，虽然 ResNet-18 网络的层数较少，训练所需的时间也较少，但其实现对胸片的分类的正确率并不低于其余较为复杂的网络。衡量网络好坏的标准是整个网络的 AUROC (Area Under the Receiver Operator Characteristic)。由于 baseline 中的网络的输出为各类疾病的患病概率，因此在没有设置阈值对病进行判断时，AUROC 可以较为全面地衡量一个分类器的性能。我们将 baseline 复现，得到的结果如下：

分类结果	ResNet18	DenseNet121	ResNet50
AUROC mean	0.8165	0.7934	0.8099
Atelectasis (肺不张)	0.7772	0.7549	0.7679
Cardiomegaly (心脏肥大)	0.8808	0.8772	0.8814
Effusion (胸膜积液)	0.8299	0.8180	0.8323

Infiltration (嗜酸粒细胞性肺浸润)	0.7009	0.6837	0.7027
Mass (肺肿块)	0.8324	0.8096	0.8244
Nodule (肺结节)	0.7719	0.7282	0.7532
Pneumonia (肺炎)	0.7306	0.7041	0.7287
Pneumothorax (气胸)	0.8553	0.8371	0.8412
Consolidation (肺实变)	0.7542	0.7450	0.7512
Edema (肺水肿)	0.8495	0.8380	0.8478
Emphysema (肺气肿)	0.9040	0.8474	0.9019
Fibrosis (肺纤维化)	0.8325	0.8197	0.8243
Pleural_Thickening (胸膜增厚)	0.7820	0.7601	0.7689
Hernia (肺疝)	0.9307	0.8847	0.9136

四、使用 Deep Mutual Learning 重新训练

1. 蒸馏学习介绍 (Model Distillation)

蒸馏学习是一种用复杂模型训练简单模型的训练方法,可以用于完成复杂神经网络的轻量化部署。蒸馏学习首先要有一个已经训练好的复杂的网络 (Cumbersome Model) 和一个还未训练的简单的, 参数较少且结构较为相似的网络 (Distilled Model), 两个网络需要完成的分类任务相同, 或者是简单网络需要完成的任务是复杂网络需要完成任务的一个子集。蒸馏学习的目的是利用已有的复杂网络来加速简单网络的训练。其做法是将训练样本同时输入到两个网络中, 利用两个网络 softmax 层的交叉熵 (Cross-entropy) 来对简单的网络进行训练。由于 softmax 层为分类器得出的各类的概率, 其比输出层所包含的信息更为丰富, 因此用交叉熵作为损失函数利用了更多的信息, 因此训练速度要较直接利用输出层的误差进行反向求导要快得多。所以使用蒸馏学习来训练简单模型时, 学习速率可以更高, 训练数据可以更少。

下面给出交叉熵 (Cross-entropy) 的定义:

	Logits	Probability (SoftMax)
Cumbersome Model	v_i	$p_i = \frac{\exp\left(\frac{v_i}{T}\right)}{\sum \exp\left(\frac{v_j}{T}\right)}$
Distilled Model	z_i	$q_i = \frac{\exp\left(\frac{z_i}{T}\right)}{\sum \exp\left(\frac{z_j}{T}\right)}$

Loss 函数为交叉熵 (Cross-entropy):

$$L_c = - \sum p_i \ln(q_i)$$

2.Deep Mutual Learning 介绍

相互学习 (Deep Mutual Learning) 是蒸馏学习 (Model Distillation) 的延展, 其用于两个不同的模型, 将两个模型同时进行训练, 利用 Softmax 层的冗余信息, 从而提高训练速率和降低误差。Deep Mutual Learning 需要两个神经网络完成的分类任务相同, 训练样本相同。其训练方法为先对一个网络进行一次训练, 待训练误差回传更新完权值后, 开始对另一个网络进行一次训练, 同样待训练误差回传更新完权值后, 再训练第一个网络, 二者交替训练。为了使得后一个网络在训练时用到前一个网络训练时已经学习到的信息, 损失函数定义为标签和 softmax 层的交叉熵与两个网络 KL 散度之和。

训练样本 $X = \{x_i\}_{i=1}^N$

	Logits	Probability (SoftMax)
Network θ_1	z_{1m}	$p_{1m} = \frac{\exp(z_{1m})}{\sum \exp(z_{1j})}$
Network θ_2	z_{2m}	$p_{2m} = \frac{\exp(z_{2m})}{\sum \exp(z_{2j})}$

Loss 函数为交叉熵 (Cross-entropy):

$$L_{c1} = \sum_{i=1}^N \sum_{m=1}^M I(y_i, m) \ln(p_{1m}(x_i))$$

其中,

$$I(y_i, m) = \begin{cases} 1 & y_i = m \\ 0 & y_i \neq m \end{cases}$$

$p_{1m}(x_i)$ 为输入数据为 x_i 时预测第 m 类的概率, y_i 为实际分类的结果。

由于我们的问题是一个多分类的问题,

KL 散度 (KL Divergence, Kullback-Leibler Divergence)

$$D_{KL}(\vec{p}_2 \parallel \vec{p}_1) = \sum_{i=1}^N \sum_{m=1}^M p_{1m}(x_i) \ln \frac{p_{2m}(x_i)}{p_{1m}(x_i)}$$

$$L_{\partial_1} = L_{c1} + D_{KL}(\vec{p}_2 \parallel \vec{p}_1)$$

类似的,

$$L_2 = L_{c2} + D_{KL}(\vec{p}_1 \parallel \vec{p}_2)$$

Algorithm:

Input:	Training set X, label set Y
	learning rate γ_{1t} and γ_{2t}
Initialize:	∂_1, ∂_2 (different initial condition of two models)
Repeat:	t=t+1
	Randomly sample data from X
	Update the predictions p_1 and p_2
	Compute the stochastic gradient and update ∂_1 :
	$\partial_1 \leftarrow \partial_1 + \gamma_{1t} \frac{\partial L_{\partial_1}}{\partial \partial_1}$
	Update the predictions p_1 and p_2
	Compute the stochastic gradient and update ∂_2 :
	$\partial_2 \leftarrow \partial_2 + \gamma_{2t} \frac{\partial L_{\partial_2}}{\partial \partial_2}$
Until:	Convergence

3.训练结果

分类结果	Res-net18	Dense-net121	Resnet18-Resnet50		Resnet18-Resnet18-	
			Res-net50	Res-net18	Res-net18	Res-net18
AUROC mean	0.810	0.793	0.811	0.791	0.808	0.800
Atelectasis	0.768	0.755	0.779	0.760	0.765	0.761
Cardiomegaly	0.881	0.877	0.874	0.861	0.878	0.877
Effusion	0.832	0.818	0.826	0.812	0.828	0.825
Infiltration	0.703	0.684	0.709	0.703	0.701	0.699
Mass	0.824	0.801	0.813	0.785	0.829	0.819
Nodule	0.753	0.728	0.765	0.732	0.758	0.740
Pneumonia	0.729	0.704	0.735	0.714	0.717	0.710
Pneumothorax	0.841	0.837	0.864	0.849	0.857	0.853
Consolidation	0.751	0.745	0.754	0.738	0.752	0.749
Edema	0.847	0.838	0.848	0.831	0.837	0.833
Emphysema	0.902	0.847	0.911	0.872	0.896	0.884
Fibrosis	0.824	0.820	0.814	0.804	0.822	0.809
Pleural	0.768	0.760	0.771	0.755	0.767	0.757

Thickening						
Hernia	0.913	0.885	0.887	0.854	0.908	0.882

关于收敛速度：

	单独训练		Deep Mutual Learning		Deep Mutual Learning	
	Resnet18	Densenet121	Resnet50	Resnet18	Resnet18	Resnet18
迭代次数	50	50	30		30	
收敛次数	18	49	8	3	10	10

这里的收敛次数是指之后几乎没有出现 Loss 小于当前 Loss 的情况，分别对应上一个表格中的训练结果。我们能可以看出，Densenet121 的收敛速率较慢，直到训练到 50ci 时还训练 2-3 次会出现一次 Loss 更小的情况。而 Resnet18 由于其网络层数较少，在单独训练时收敛速度较 Densenet121 快。而在 Deep Mutual Learning 中，由于是两个网络交替训练，其训练时可以用到另一个网络已经学习到的部分信息，因此其每个网络所需的训练次数大幅度减少，但训练次数之和还是要高于一个网络单独训练。正确率上而言，几乎与一个网络单独训练持平。

五、分类训练

1.病理特征的介绍

肺病种类	病理特征
Atelectasis (肺不张)	不张的肺组织透亮度降低，均匀性密度增高，恢复期或伴有支气管扩张时可密度不均（囊状透亮区）。不同程度的体积缩小，亚段及以下的肺不张可因有其他侧枝的通气而体积缩小不明显。叶段性肺不张一般呈钝三角形，宽而纯的面朝向肋膈胸膜面，尖端指向肺门，有扇形、三角形、带状、圆形等。叶间裂向不张的肺侧移位，如右肺横裂叶间胸膜移位，两侧的斜裂叶间胸膜移位等；由于肺体积缩小，病变区的支气管与血管纹理聚拢，而邻近肺代偿性膨胀，指使血管纹理稀疏，并向不张的肺叶弓形移位；肺门阴影向不张的肺叶移位；肺门阴影缩小和消失，并且与肺不张的致密影相隔合；纵隔、心脏、气管向患侧移位，特别是全肺不张时明显，有时健侧肺疝移向患侧，而出现纵隔疝；横膈肌升高，胸廓缩小，肋间隙变窄。
Cardiomegaly (心脏肥大)	心脏所占亮区比例过大
Effusion (胸膜积液)	在标准的后胸部 x 光片上，胸膜积液呈现为白色区域。正常情况下，脏胸膜和胸膜壁层之间的空间是看不见的。胸膜积液渗透到这些层之间的空隙。因为胸膜积液的密度与水相似，可以在放射

	学上看到。由于积液比肺的其他部分密度更大，所以它会倾向于胸膜腔的下部。胸膜积液表现为基本的流体动力学，符合胸膜腔的形状。
Infiltration（嗜酸粒细胞性肺浸润）	X 线表现常为密度较淡、边界不清的片状阴影，分布于单侧或双侧肺部，呈短暂游走性。
Mass（肺肿块）	肺良性肿瘤多有包膜，呈边缘锐利光滑的球形肿块，生长慢，一般不发生坏死。肺含液囊肿的 X 线表现与良性肿瘤不易区分，但含液囊肿可随深呼吸运动而有形态的改变。恶性肿瘤多无包膜，呈浸润性生长，故边缘多不锐利，并可有短细毛刺伸出。由于生长不均衡，其轮廓常呈分叶状或有脐样切迹。生长快，可发生中心坏死。
Nodule（肺结节）	常表现为肺门增大，以双侧肺门肿大为主，易误诊为肺结核或肿瘤
Pneumonia（肺炎）	细菌感染的社区性肺炎典型症状有一个肺叶节段的肺实变，称为大叶性肺炎。然而并非所有的细菌性肺炎皆为大叶性肺炎，也并非所有的大叶性肺炎皆由细菌所引起，其他肺炎也会有类似的情形。吸入性肺炎的特点是肺两侧底部及右肺叶呈现混浊。病毒性肺炎的 X 光片可能是正常的，也可能有过度充气、两侧大片浸润、或是类似细菌性肺炎的肺实变情形。
Pneumothorax（气胸）	气胸胸片上大多有明确的气胸线，即萎缩肺组织与胸膜腔内的气体交界线，呈外凸线条影，气胸线外为无肺纹理的透光区，线内为压缩的肺组织。大量气胸时可见纵隔、心脏向健侧移位。合并胸腔积液时可见气液面。局限性气胸在后前位 X 线检查时易漏诊，侧位胸片可协助诊断，X 线透视下转动体位也可发现。若围绕心缘旁有透光带应考虑有纵隔气肿。
Consolidation（肺实变）	肺实变的患者中，充气的支气管在平片或 CT 上都能看到，在不透光的肺实质衬托下显示为透光区。
Edema（肺水肿）	肺泡水肿主要表现为腺泡状致密阴影，呈不规则相互融合的模糊阴影，弥漫分布或局限于一侧或一叶，或从肺门两侧向外扩展逐渐变淡成典型的蝴蝶状阴影。有时可伴少量胸腔积液。
Emphysema（肺气肿）	胸廓扩张，肋间隙增宽，肋骨平行，膈降低且变平，两肺野透亮度增加。
Fibrosis（肺纤维化）	早期虽有呼吸困难，但 X 线胸片可能基本正常；中后期出现两肺中下野弥散性网状或结节状阴影，偶见胸膜腔积液，增厚或钙化。
Pleural Thickening（胸膜增厚）	<p>胸膜粘连：指两层的胸膜粘着在一起。这种病变是由肺结核、胸膜炎以及胸部损伤后引发的。</p> <p>胸膜转移：主要来源于肺，其次是乳腺，其他常见原发部位有胃，卵巢，胰腺等。癌症造成胸膜层毛细血管内压，胶体渗透压，毛细血管通透性和胸腔内压力改变，产生胸腔积液——恶性胸水。恶性胸水又称癌性胸膜炎，恶性胸腔积液是晚期癌症的常见并发</p>

	症，临床上癌症患者一旦出现胸腔积液，即意味着病变已局部或周身播散。
	胸膜钙化：胸膜腔内有机化的血块或干酪坏死物质等存在时可有钙盐沉着，形成胸膜钙化。胸膜钙化多见于结核性胸膜炎、化脓性胸膜炎及损伤性血胸后。某些尘肺，如滑石肺及石棉肺也可有胸膜钙化，而且常呈双侧性的。胸膜钙化常和胸膜增厚和粘连同时存在。
Hernia（肺疝）	颈部肺疝 X 线可见颈部含气的软组织块影与肺尖部的肺组织相连。颈部软组织块影的大小与胸腔内压力呈明显正相关，吸气时软组织影增大，呼气时缩小，呈规律的周期性增大缩小改变。
	胸部肺疝切线位摄胸片 可见到透光度增强的膨出肿块，与肺组织相连。

2.分类训练的结果

由于各种病的特征不尽相同，有些疾病在胸片中表现为肺部轮廓发生了显著的变化，如胸膜增厚，心脏肥大等，而有些病症在胸片中的表现则是细节上的变化，如出现阴影，纹理发生改变，如肺炎等。由于卷积神经网络卷积核的变化可以使得网络对于图片关注的层次发生一定的改变，如卷积核越大其对全局信息的感受能力越强。基于此思路，我们按照各种病症特征的不同对特征进行 4 分类，分别训练了 4 个分类器，来完成对胸片病症的判断。每个分类器只负责其所辖疾病概率的输出。

第一类	第二类	第三类	第四类
肺不张 胸膜积液 嗜酸粒细胞性肺浸润 气胸 肺水肿 胸膜增厚	肺肿块 肺结节 肺实变 肺疝	心脏肥大 肺气肿	肺炎 肺纤维化

分类结果	ResNet18	分 4 类结果
AUROC mean	0.8165	0.8154
Atelectasis（肺不张）	0.7772	0.7723
Cardiomegaly（心脏肥大）	0.8808	0.8869
Effusion（胸膜积液）	0.8299	0.8233
Infiltration（嗜酸粒细胞性肺浸润）	0.7009	0.7020
Mass（肺肿块）	0.8324	0.8277
Nodule（肺结节）	0.7719	0.7783
Pneumonia（肺炎）	0.7306	0.7216

Pneumothorax (气胸)	0.8553	0.8565
Consolidation (肺实变)	0.7542	0.7395
Edema (肺水肿)	0.8495	0.8481
Emphysema (肺气肿)	0.9040	0.9245
Fibrosis (肺纤维化)	0.8325	0.8176
Pleural_Thickening (胸膜增厚)	0.7820	0.7805
Hernia (肺疝)	0.9307	0.9373

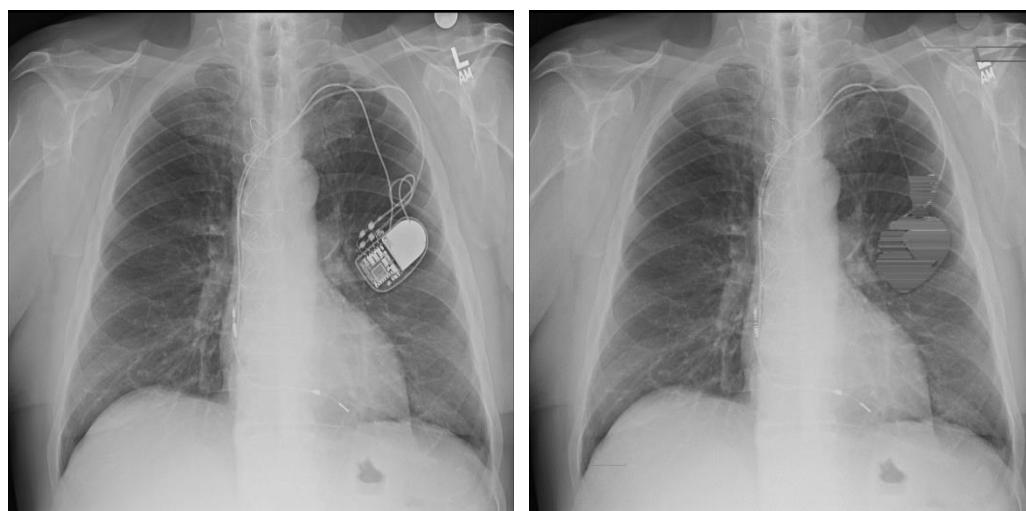
六、图像的预处理

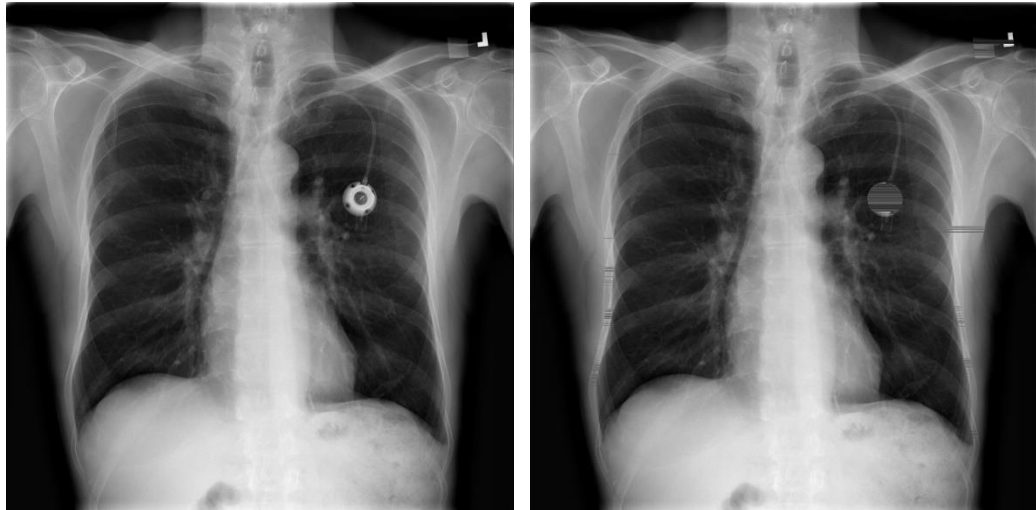
考虑到原胸片数据库中有相当比例的图片上包含了心脏起搏器、线管等高亮图形，而这些图形毫无疑问会对识别系统造成一些负面干扰，特别是在考虑到神经网络对图像做卷积运算时可能因为这些高亮度的干扰而产生较大误差，所以我们在输入这些图片之前对它们进行预处理。

而由于数据量较大，我们采用了比较简单直接的逐行扫描并操作的方式。具体来讲，当某一行的像素值在一个位置发生向上的突变（相邻两个像素点差值达到了事先设置的像素差阈值），并且仅在较少的宽度下保持了高亮度后接着发生向下的突变，那么我们认为这些高亮度图形为干扰，从而用两端的较暗像素值的平均值替换填充这些位置。

在经过一系列实验后，我们发现将像素差阈值设置为 50，将宽度阈值设为 40 左右时能达到较好的处理效果。

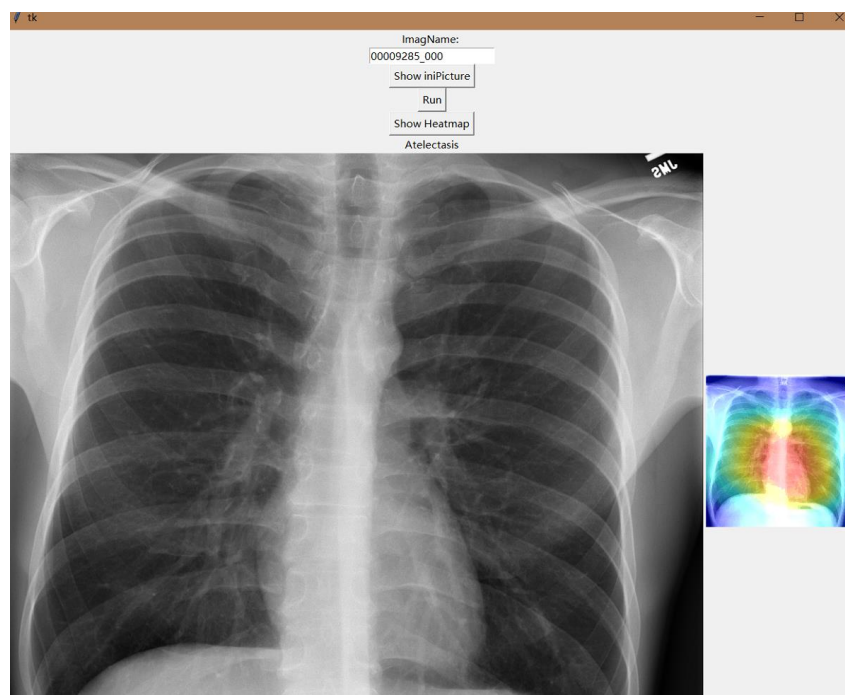
处理前后的图片对比：





七、可视化展示

我们还基于python中的tkinter完成了图形界面的制作,其输入一张胸片,可以将可能患病的区域进行标注。



八、实验总结

我们在这个实验中收获了很多，首先这是一个开放的问题，我们需要自己发现问题所在并解决问题，我们在不断发现问题不断尝试的过程中得到了锻炼。当然，在运用各种方法的过程中，我们更加深刻地体会了课内所讲到的神经网络有关知识，也发现了实际运用和理论情况间存在的差距，不断克服大大小小的困难是我们最终能够获得目前成果的必经之路。

这次的实验结果并不算理想，能改进的地方还有很多，如多分类时可以训练一个大的分类器，以避免较多不相关的样本对四个模型的影响。多分类是根据我们调研医学文献再加上自己的判断来进行的分类，并没有用到其中的样本中的信息。还可以在对病症进行分类前对样本采用聚类等算法进行预处理，也许会对分类有所帮助。

九、参考文献

- [1] Pranav Rajpurkar, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, Aarti Bagul CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning
- [2] Ying Zhang, Tao Xiang, Timothy M. Hospedales, Huchuan Lu Deep Mutual Learning
- [3] Geoffrey Hinton, Oriol Vinyals, Jeff Dean Distilling the Knowledge in a Neural Network
- [4] ChestX-ray8: Hospital-scale ChestX-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases
- [5] Gao Huang, Zhuang Liu, Laurens van der Maaten Densely Connected Convolutional Networks