



東南大學

本科毕业设计（论文）开题报告

毕业设计	基于眼底图像的多任务
（论文）题目	诊断模型的设计与实现

学号：	06219109
-----	----------

姓名：	孙寒石
-----	-----

学院：	电子科学与工程学院
-----	-----------

专业：	电子科学与技术
-----	---------

指导教师：	周毅
-------	----

开题日期：	2023-02-20
-------	------------

一、选题背景和意义：

随着社会的不断发展和科技的不断进步，人们对于医疗卫生的需求越来越高。而且，随着人口老龄化和慢性病的不断增加，对医疗卫生服务的需求将会进一步增加。在这种情况下，如何提高医疗卫生服务的质量和效率成为了一个非常重要的问题。医疗卫生服务质量和效率的提高，涉及到医疗资源的合理配置、医疗流程的优化、医疗信息化建设等多个方面。在医疗信息化建设中，医疗图像处理技术是一个非常重要的研究方向。医疗图像处理技术可以对医疗影像进行处理和分析，提高医生对医疗影像的判断和诊断能力。尤其是在医疗影像诊断中，医疗图像处理技术可以对医疗影像进行增强、降噪、分割、配准、重建等处理，从而提高医生对医疗影像的诊断准确性和可靠性。眼科学中使用的眼底彩照图像(CFP)等影像资源丰富，但是医学影像 AI 领域面临的数据集小而杂的问题同样存在。这是因为不同医疗机构采用的影像设备和影像采集方式不同，导致生成的影像数据具有很大的差异性。此外，医学影像数据涉及患者隐私，因此很难收集足够多的公开数据集来支持广泛的研究。

针对这个问题，一些研究人员采用迁移学习的方法来解决数据集小而杂的问题。迁移学习是指将一个已经训练好的模型应用于一个新的任务，而不是从头开始训练一个新模型。在眼底图像诊断领域，研究人员可以使用一个在大规模数据集上预训练好的神经网络模型，然后针对特定任务进行微调，以适应眼底图像数据的特点。这种方法可以显著提高模型的泛化能力和诊断准确率，同时节省大量的训练时间和计算资源。我们的任务旨在利用眼底图像和人工智能技术来辅助识别多种疾病，具有非常广阔的应用前景。多任务学习算法和多源域迁移算法可以帮助解决数据集小而杂的问题，提高模型的泛化能力和迁移性能，从而使得模型具有更好的适应性和可扩展性。在多个不同任务的公开数据集上进行研究与实验，也有助于更全面地评估模型的性能和鲁棒性。同时，面对下游新数据集可以高效地进行知识迁移，无需从头学习。这不但节约了模型所占用的资源，而且最大化了对多个不同数据集的利用。

二、课题关键问题及难点：

本课题需要设计实现一个多疾病辅助识别模型，需要同时处理多个任务，包括分类和分割等。如何在不同任务之间平衡模型的学习和避免任务之间的干扰，同时提高模型的精度和效率是一个关键问题。

- 1. 类别不平衡问题：**在眼底图像中，各类别的数量存在不平衡的现象，一些疾病较为罕见，而一些常见疾病的样本数量又非常庞大，这将影响模型的性能表现。
- 2. 不同数据集之间的差异性：**在眼底图像的研究中，存在许多不同来源的数据集，这些数据集可能存在不同的病变类型、不同的拍摄设备、不同的标注标准等，这些差异会对模型的泛化性能造成影响。
- 3. 疾病之间的相似性和差异性：**在多疾病识别任务中，不同疾病之间存在相似和差异，模型需要能够准确区分不同疾病，并在相似疾病之间做出准确的选择。
- 4. 多疾病协同识别问题：**不同疾病之间存在相互影响的情况，一个疾病的出现可能预示着另一个疾病的风险增加，模型需要考虑到这些关联关系，做到多疾病协同识别。
- 5. 模型泛化能力问题：**在实际应用中，模型需要具有较好的泛化能力，能够适应不同来源的数据集和不同的临床环境，而这需要模型在训练时具有良好的泛化能力。

注：开题报告可单独装订，但在院（系）范围内，封面和装订格式必须统一。

三、文献综述（或调研报告）：

近年来，随着人工智能技术的发展，眼底图像在医疗领域中的应用越来越广泛 [16, 17]。眼底图像可以通过计算机视觉和机器学习技术进行自动分析和诊断，使得医生能够更加准确地判断病情，提高了诊断的效率和精度 [18, 19]。本文将综述当前眼底图像多疾病诊断的相关研究。首先，多种神经网络模型被应用于眼底图像诊断。卷积神经网络（CNN）是最常见的一种神经网络，它可以自动从图像中提取特征并进行分类。CNN 在眼底图像的分类、疾病检测和分割等任务中得到广泛应用。例如使用 ResNet-50 模型对眼底图像进行分类和疾病检测，取得了很好的效果 [22, 23]。另外，Densenet、Inception 等模型也被用于眼底图像的多疾病分类 [24, 25]。

迁移学习被广泛应用于眼底图像多疾病诊断。由于医学图像数据集通常较小，因此在训练 CNN 模型时容易出现过拟合的问题 [20, 21, 28]。迁移学习可以解决这个问题，通过在大规模图像数据集上训练好的模型参数作为初始化参数，迁移到小规模医学图像数据集上进行微调。迁移学习不仅可以提高模型的准确率，而且可以缩短模型的训练时间。由于不同数据集之间的分布差异，模型在不同数据集上的表现往往存在较大差异。传统的做法是在新数据集上重新训练模型，但这样会浪费大量时间和资源 [26, 27, 28]。而迁移学习则可以在已有的数据集上训练出一个泛化能力较强的模型，并通过微调或其他方式将其应用于新的数据集，从而减少训练时间和提高模型性能 [29]。一些研究者也尝试将多源迁移学习技术应用于眼底图像分类问题，尝试在不同数据集之间共享知识以提高模型的泛化能力 [29, 30]。

多任务学习 [10, 11, 12] 是指在一个模型中学习多个相关任务的能力。相较于单独训练多个独立的模型，多任务学习可以提高模型的泛化能力、减少计算资源的消耗，同时还可以帮助解决数据稀疏性问题 [13, 14, 15]。在过去几年中，许多多任务学习的算法被提出，包括模型架构上的创新和梯度反传上的优化等。Hard Parameters Sharing [1] 算法是一种经典的多任务学习算法。该算法通过在不同任务之间共享模型的参数来学习多个任务。这种共享参数的方式被称为“硬共享”，因为不同任务之间共享的参数是固定的。这种方法可以减少参数数量，同时提高模型的泛化能力。但是由于所有参数都被共享，所以模型的表示能力受到限制，不能很好地处理任务之间的差异和共性。MMoE（Mixture of Experts）[2] 算法通过将不同的任务分配给不同的专家网络，然后将它们的结果混合在一起来实现任务之间的共享。MMoE 模型中每个专家网络都针对特定的任务进行优化，并且通过学习每个任务的专家网络之间的关系来提高模型的整体性能。MTAN（Multi-Task Attentional Network）[3] 算法是一种基于注意力机制的多任务学习算法。MTAN 模型中，每个任务都有自己的注意力机制，用于选择不同的特征子集。这些特征子集被用来进行任务特定的处理，然后将结果进行聚合。MTAN 通过在不同的任务之间共享底层特征，同时学习每个任务的特定处理。Residual Adapters [4] 算法是一种适用于多任务学习的改进的 ResNet 模型。该算法通过在 ResNet 模型的不同层中添加一些适配器来实现不同的任务特定的处理。这些适配器能够在不影响整个模型性能的情况下，针对不同的任务进行微调。

Conflict-Averse GD（Gradient Descent）[5] 算法是一种用于多任务学习的优化算法。它通过在目标函数中引入一个“冲突惩罚项”，来减少任务之间的冲突。该算法的优化目标是 최소화冲突惩罚项和所有任务的误差的加权和。Gradient Vaccine [6] 算法是一种通过向模型注入类似疫苗的正则化项，来改善多任务学习模型的泛化能力。这种正则化项会防止模型过拟合每个任务，从而增强模型的泛化性能。Fishr [7] 算法是一种用于领域泛化的多任务学习算法。Fishr 通过利用训练数据中的共性和差异来学习一个领域通用的模型，从而实现跨领域的

泛化能力。Task Grouping [8] 算法是一种用于多任务学习的任务聚类算法。该算法将不同的任务聚类为不同的任务组，然后在每个任务组中训练一个多任务学习模型。这种方法可以减少任务之间的干扰。DoDNet [9] 中引入了一个动态的滤波器来作为实现动态输出，从而根据输入数据动态地生成一个网络来处理。

参考文献:

[1] Ruder, Sebastian. "An overview of multi-task learning in deep neural networks." arXiv preprint arXiv:1706.05098 (2017).

[2] Ma, Jiaqi, et al. "Modeling task relationships in multi-task learning with multi-gate mixture-of-experts." Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining. 2018.

[3] Wang, Cheng, et al. "Manacs: A multi-task attentional network with curriculum sampling for person re-identification." Proceedings of the European conference on computer vision (ECCV). 2018

[4] Rebuffi, Sylvestre-Alvise, Hakan Bilen, and Andrea Vedaldi. "Learning multiple visual domains with residual adapters." Advances in neural information processing systems 30 (2017).

[5] Liu, Bo, et al. "Conflict-averse gradient descent for multi-task learning." Advances in Neural Information Processing Systems 34 (2021): 18878-18890.

[6] Wang, Zirui, et al. "Gradient vaccine: Investigating and improving multi-task optimization in massively multilingual models." arXiv preprint arXiv:2010.05874 (2020).

[7] Rame, Alexandre, Corentin Dancette, and Matthieu Cord. "Fishr: Invariant gradient variances for out-of-distribution generalization." International Conference on Machine Learning. PMLR, 2022.

[8] Fifty, Chris, et al. "Efficiently identifying task groupings for multi-task learning." Advances in Neural Information Processing Systems 34 (2021): 27503-27516.

[9] Zhang, Jianpeng, et al. "Dodnet: Learning to segment multi-organ and tumors from multiple partially labeled datasets." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021.

[10] Rich Caruana. Multitask learning. Machine learning, 1997.

[11] Davy Neven, Bert De Brabandere, Stamatios Georgoulis, Marc Proesmans, and Luc Van Gool. Fast scene understanding for autonomous driving. In IV Workshops, 2017.

[12] Alex Kendall, Yarin Gal, and Roberto Cipolla. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In CVPR, 2018.

[13] Zhao Chen, Vijay Badrinarayanan, Chen-Yu Lee, and Andrew Rabinovich. Gradnorm: Gradient normalization for adaptive loss balancing in deep multitask networks. In ICML, 2018.

[14] Ozan Sener and Vladlen Koltun. Multi-task learning as multi-objective optimization. In NIPS, 2018.

[15] Marvin Teichmann, Michael Weber, Marius Zoellner, Roberto Cipolla, and Raquel Urtasun. Multinet: Real-time joint semantic reasoning for autonomous driving. In IV, 2018.

[16] Gulshan, Varun, et al. "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs." Jama 316.22 (2016): 2402-2410.

[17] Ting, Daniel Shu Wei, et al. "Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes." Jama 318.22 (2017): 2211-2223.

- [18] Rajalakshmi, Ramachandran, et al. "Automated diabetic retinopathy detection in smartphone-based fundus photography using artificial intelligence." *Eye* 32.6 (2018): 1138-1144.
- [19] Burlina, Philippe M., et al. "Automated grading of age-related macular degeneration from color fundus images using deep convolutional neural networks." *JAMA ophthalmology* 135.11 (2017): 1170-1176.
- [20] Ting, Daniel SW, et al. "AI for medical imaging goes deep." *Nature medicine* 24.5 (2018): 539-540.
- [21] Gargeya, Rishab, and Theodore Leng. "Automated identification of diabetic retinopathy using deep learning." *Ophthalmology* 124.7 (2017): 962-969.
- [22] Brown, James M., et al. "Automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks." *JAMA ophthalmology* 136.7 (2018): 803-810.
- [23] Treder, Maximilian, Jost Lennart Lauermann, and Nicole Eter. "Deep learning-based detection and classification of geographic atrophy using a deep convolutional neural network classifier." *Graefe's Archive for Clinical and Experimental Ophthalmology* 256 (2018): 2053-2060.
- [24] Li, Ning, et al. "A benchmark of ocular disease intelligent recognition: One shot for multi-disease detection." *Benchmarking, Measuring, and Optimizing: Third BenchCouncil International Symposium, Bench 2020, Virtual Event, November 15–16, 2020, Revised Selected Papers 3*. Springer International Publishing, 2021.
- [25] Yongxi Lu, Abhishek Kumar, Shuangfei Zhai, Yu Cheng, Tara Javidi, and Rogerio Feris. Fullyadaptive feature sharing in multi-task networks with applications in person attribute classification. In *CVPR*, 2017.
- [26] Abramoff, Michael David, et al. "Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning." *Investigative ophthalmology & visual science* 57.13 (2016): 5200-5206.
- [27] Azzopardi, George, et al. "Trainable COSFIRE filters for vessel delineation with application to retinal images." *Medical image analysis* 19.1 (2015): 46-57.
- [28] Kandel, Ibrahem, and Mauro Castelli. "Transfer learning with convolutional neural networks for diabetic retinopathy image classification. A review." *Applied Sciences* 10.6 (2020): 2021.
- [29] Bhardwaj, Charu, Shruti Jain, and Meenakshi Sood. "Transfer learning based robust automatic detection system for diabetic retinopathy grading." *Neural Computing and Applications* 33.20 (2021): 13999-14019.
- [30] Burlina, Philippe, et al. "Comparing humans and deep learning performance for grading AMD: a study in using universal deep features and transfer learning for automated AMD analysis." *Computers in biology and medicine* 82 (2017): 80-86.

四、方案（设计方案、或研究方案、研制方案）论证：

本课题需要研究的问题：

该课题需要综合运用多种技术和算法，并结合多个不同任务的眼底公开数据集开展研究和实验，从而实现一个高效、准确的眼底图像辅助诊断系统，同时需要解决眼底图像领域的数据集小而杂的问题。眼科学对影像学检查依赖性强，影像资源丰富，但不同数据集的来源、疾病类别和任务类型却各不相同。因此，需要研究多任务学习算法和多源域迁移算法，以使不同任务数据训练互相促进，训练所得模型通用性好，面对下游新数据集的到来，可以高效地进行知识迁移，无需从头学习。

拟采用的方法：

1. 数据集的选取：选取十个分类数据集，包括单分类、二分类和多分类。

我选取了 TAOP, APTOS, DDR; AMD, LAG, PALM, REFUGE; ODIR-5K, RFMiD, KaggleDR+这十个数据集。选取的这十个眼底数据集在眼底图像研究领域中都是比较有代表性的公开数据集，涵盖了不同的疾病类型、图像来源、图像质量等方面，具有以下一些优势：

- **多样性：**您选取的这些数据集覆盖了多种不同的眼底疾病类型，包括老年性黄斑变性、青光眼、糖尿病视网膜病变等等。这些数据集中的图像来源也多样，包括来自医院、诊所设备等，具有一定的代表性和普适性。
- **数据规模：**选取的一些数据集，如 KaggleDR+、ODIR-5K 等，拥有大量的图像样本，可用于训练较为复杂的深度学习模型，并且这些数据集中的图像质量相对较高，可以提高模型的诊断准确性。
- **真实性：**这些数据集的图像均来自于真实临床检查或者诊断，具有较高的真实性和临床可应用性，可以用于验证算法在实际临床环境中的表现和效果。
- **已有研究：**这些数据集已经被广泛应用于眼底图像处理和疾病诊断方面的研究中，并且已经有一定的标注和基准模型可供参考，可以为您的实验提供参考和比较。

所以选取这些眼底数据集可以更好地探索不同的疾病分类、图像来源、图像质量等问题，拓展实验的视野和深度，同时也可以为眼底图像处理和疾病诊断领域的研究提供更多的数据支持和验证。

2. 数据预处理：采用 CLAHE 对所有眼底图像进行预处理，并将图片 resize 到 256*256。

CLAHE（Contrast Limited Adaptive Histogram Equalization）是一种常用的图像增强算法，可以有效地改善图像的对比度和细节，特别适用于医学图像处理。在眼底图像处理中，CLAHE 算法通过分割图像为小块，对每个块进行直方图均衡化，并限制均衡化后像素灰度级别的最大值，从而避免对图像整体对比度的破坏，保留图像的细节信息，提高图像的视觉质量和诊断准确性。与传统的直方图均衡化算法相比，CLAHE 算法可以更好地处理图像中的局部对比度，对于眼底图像中视网膜血管、黄斑、视神经等重要部位的显示具有更好的效果。

需要注意的是，CLAHE 算法在应用时需要调节一些参数，如窗口大小、块大小、直方图均衡化程度等，以适应不同的图像特性和实际需求。同时，CLAHE 算法也存在一些缺陷，如可能引入噪声、产生图像假影等问题，需要在实际应用中进行综合考虑和优化。

3. 基线模型：采取 HPS、MMoE、MTAN、Gradient Vaccine、Fishr、Conflict-Averse Gradient Descent、Residual Adapters 等多种模型。

由综述可知，多任务学习方法主要包括梯度策略、新的模型框架、好的预训练权重等。我选取这些模型作为基线模型是因为：

多样性：这些模型来自于不同的研究方向和算法框架，有些是基于新的网络的，有些是基于梯度下降的，有些是结合了多个任务参数的。因此，采用这些模型可以使得研究更全面，避免单一方法的局限性。

先进性：这些模型都是近年来在相关领域中取得了一定成果的，采用这些模型可以借鉴前人的研究成果，从而快速提高研究效率和准确性。

可扩展性：这些模型都具有一定的可扩展性，可以通过调整参数或者结构来适应不同的任务和数据集。因此，采用这些模型可以为未来的研究提供一个良好的基础，同时也可以使得研究具有更广泛的适用性。

在进行实验比较和分析，寻找出最适合眼底数据场景的多任务学习方案。同时实现各个眼底数据集的单任务学习模型，用于和后面多任务学习模型结果进行比较。

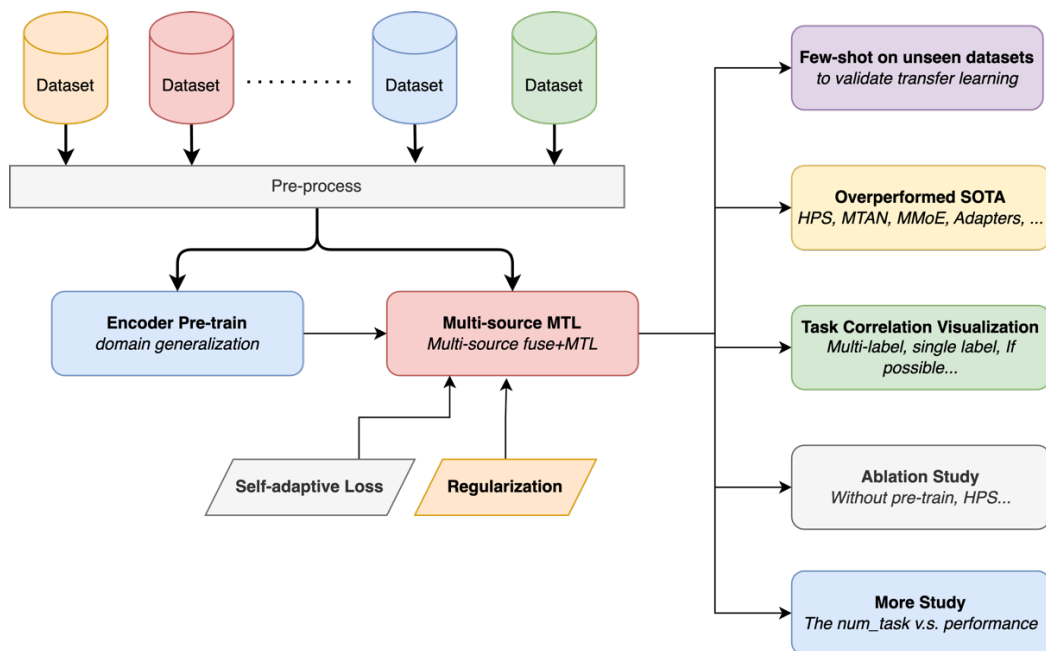
4. 改进方法

基于损失函数和梯度权重的调整：类似于 Gradient Vaccine 和 Conflict-Averse GD，创新一种自适应调整各个任务损失函数大小权重的方法，从而平衡各个任务的表现。

预训练模型的改进：实验可以得到在单任务上预先训练之后的 Encoder 要比 ImageNet 初始化权重的 Encoder 效果会有很大的提升，希望可以创新一种更好的方法得到更好的预训练权重（可以不必是和 MoCo 一样，我们可以做有监督的，效果会更好）。

模型结构的设计：考虑加入一些 Task-Specific 的参数或者加入一些别的附加网络（例如一个 Discriminator）从而提高模型的表现。

合适的正则：和 Fishr 一样，我们可以考虑设计一些正则损失函数。



5. 训练过程

数据增强：数据经过预处理后，对数据进行随机翻转、裁剪、颜色增强、Normalization 等操作后作为网络的输入。

损失函数：所有任务的损失函数经过自适应加权求和 + 一些可能的正则

优化器：Adam，初始学习率 $1e-4$ ，每个 epoch 有 200 个 batch，训练 300 个 epoch。

模型存储：每 50 个 epoch 存一次 checkpoint，同时每个 epoch 都测试一次，存下最好的模型。

评价模型的方法：

对于多任务的学习表现，我们有以下两种评价：

- **Mean Performance:** 取所有任务表现结果的均值。
- **Score:** （Single Task 为 25*10=250 分，满分 1000 分）

$$S = \sum_{d=1}^{10} \alpha_d \max\{0, \beta_d E_d^{\text{baseline}} - E_d\}^{\gamma_d} \quad (1)$$

where $E_d = 1 - \text{Score}_d$, $\alpha_d = 100\{\beta_d E_d^{\text{baseline}}\}^{-\gamma_d}$

$$S = \sum_{d=1}^{10} 100 \max\{0, 1 - E_d / \beta_d E_d^{\text{baseline}}\}^{\gamma_d} \quad (2)$$

对于新的下游任务，我们 Fix 住 Encoder，仅 Finetune 下游任务的 Task Head，衡量其表现情况。

我们还会进行一系列消融实验，包括任务数量大小对于模型的影响等。

五、进度安排：

起止日期	工作内容	备 注
2022.11	文献阅读、充分了解任务、完成单任务的基线实验。	已完成
2022.12-2023.02	完成一些 SOTA 的复现作为后续比较实验的一些基线。	已完成
2023.02	完成文献翻译与开题报告。	已完成
2023.03	算法设计及改进创新，进行实验验证；同时准备中期验收。	
2023.04	实验初具结果，进行改进，完成最后的实验和模型创新。	
2023.05	进行消融实验和一些模型评价实验，完成论文。	
2023.06	毕业答辩，代码提交验收。	