

# 基于迁移学习算法的糖尿病自动筛选系统研究

陈检 肖思隽 孙秋梅

(山东财经大学 统计学院 济南 250000)

**摘要** 现阶段我国基层医疗条件落后于城市地区，具有基层的眼科医生资源匮乏和带标记的糖尿病眼底图像数据不足等问题。针对以上问题本文提出基于迁移学习算法的糖尿病自动筛选系统，受卷积神经网络中迁移学习思想的启发，通过构造出 CNN+SVM 的融合模型来实现正确的图像分类。本次试验采用源数据来自于山东省泰安某医院。相比于直接使用 SVM 模型，CNN+SVM 融合模型训练和预测所需时间由 1993 秒下降到 1518 秒左右，kappa 值由 0.20 上升到 0.79，特异度和敏感度都提高了大概 30% 的幅度。结论：CNN+SVM 的融合模型不仅加快了模型的学习效率，而且实现了高准确率的图像分类识别。

**关键词** 糖尿病 图像分类 卷积神经网络 迁移学习 泰安某医院

## Research on automatic classification system of diabetic retinopathy based on transfer learning algorithm

CHEN Jian, XIAO Sijun, SUN Qiumei

(School of Statistics, Shandong University of Finance and Economics, Jinan250000)

### ABSTRACT

At present, the medical conditions of grassroots extremely lags behind urban areas in China. Lacking of resources for ophthalmologists with grassroots and the enough data of fundus images. In view of the above problems, this paper proposes an automatic screening system for diabetic retinopathy based on transfer learning algorithm, inspired by the idea of transfer learning in convolution neural networks, the correct image classification is achieved by constructing a fusion model of CNN + SVM. The source data used in this experiment came from a hospital in Tai'an, Shandong Province. Due to the use of the transfer learning algorithm, the time required for model training and prediction was reduced from about 1993 to about 1518 seconds, the kappa value rised from 0.20 to 0.79, and the specificity and sensitivity increased by about 30% compared with the directly use of the SVM model. The final results show that the CNN + SVM fusion model not only accelerated the learning efficiency of the model, but also achieved high accuracy image recognition and classification.

**Keywords** Diabetic Retinopathy, Image Classification, Transfer Learning, Convolution Neural Network, Tai'an Hospital

# 一、引言

## （一）研究背景和目的

### 1.研究背景

由糖尿病引起的视网膜病变称之为（Diabetic Retinopathy, DR），中国目前的眼科大夫只有 3 万多人，眼科门诊量有 8000 万人次，糖网病（DR）是造成当前人们失明的主要原因。随着人们的生活水平质量不断上升，养成了越来越多不健康的饮食习惯，导致糖尿病患者数量急速增加，一旦患上糖尿病，则有 80%

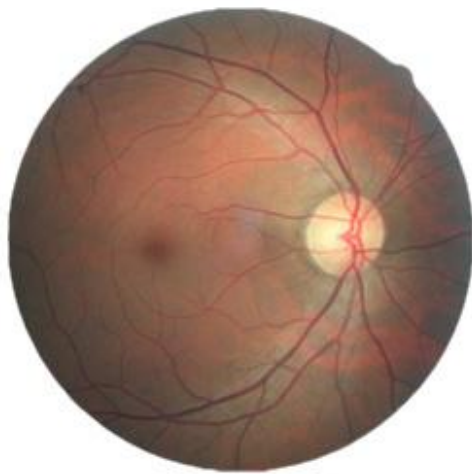


图 1 眼底图像 Fig.1 Fundus image 给出的分析结果往往再需要一两天的时间，诊断结果的延迟导致随访失效，沟通不畅和延迟治疗。

我国医院的临床诊断情况是临床医生通过观察眼底图像（左上图所示）根据经验来判断是否存在糖网病以及进一步细分症状患病的程度（之后会详细的介绍糖网病症状的分类情况）。虽然这种方法是有效的，但对于糖尿病高患病率的地区来说资源需求量巨大，这些地区最需要定期做 DR（糖网病）检测。结合我国具有显著的城乡经济二元结构的特点，农村地区和经济水平低的基层居民面临着医生缺乏临床经验和医疗设备匮乏的尴尬境地，随着糖网病患者人数持续增长，由糖网病引起的致盲症状增长率将持续上升，所需的基础设施将变得更加不足，这些问题靠传统的方法短时间内很难得到很好的解决。

### 2.研究目的

随着大数据技术的迅速发展，特别是图像识别领域，使得人们开始思考利用机器学习的思想自动对图像进行分类，从而减少医生的工作量，缓解目前医患数量严重不对称的情况。虽然从方法上看是可行的，但前提是需要大量标记好（分好类）的图像数据集作为训练模型的原始数据来训练模型。然而现阶段的情况是，很多领域很难短时间内收集到几百万张或者上千万张已经人工分好类的数据集，且花费的人力和财力成本很高，这种现象在欠发达地区尤其明显。本论文利用深度学习中迁移学习的思想，建立卷积神经网络模型，用上百万量级别的非糖网病

图像数据集来训练该模型，然后保存好该模型的结构参数，之后将少量的糖网病图像数据集导入该模型，这样做是可以将之前非糖网病的数据集中的结构参数保留下来为少量糖网病数据集训练模型所用，基于此方法可以加快模型收敛速度，显著地减少模型再次训练的所需时间，也进一步验证了不同领域之间模型的泛化能力的提升和利用少量数据也能完成图像识别分类的目的。最终的结果是得到一套高准确率的糖网病自动筛选系统,本次试验用于运行程序的主机配置为（CPU:i3-6100,3.7GHZ， GPU:GTX950M 4G）。

## （二）研究现状

糖网病诊断主要是根据是否出现微动脉瘤、渗出物、小血点、棉絮斑来描述糖网病患者患病的严重程度，如下图所示<sup>[2]</sup>：

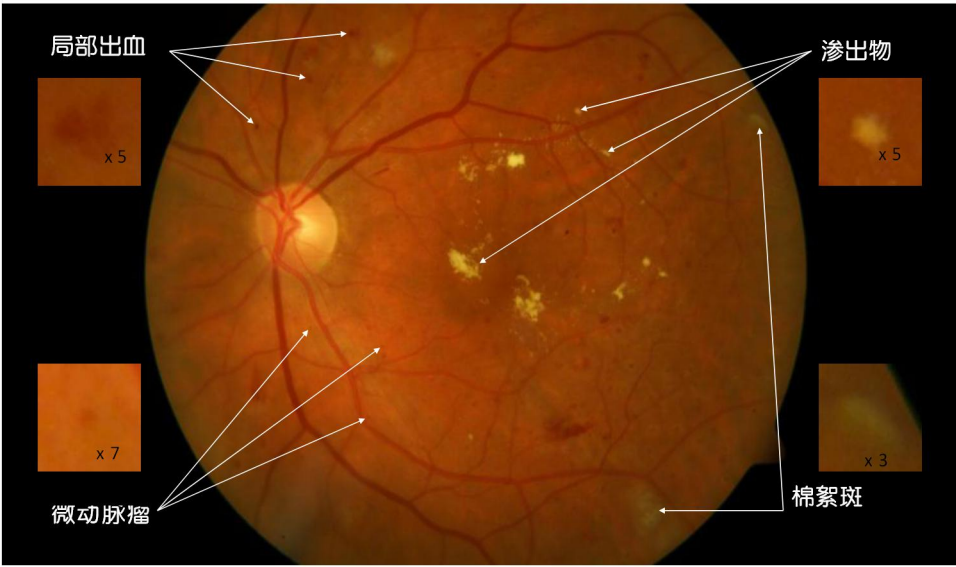
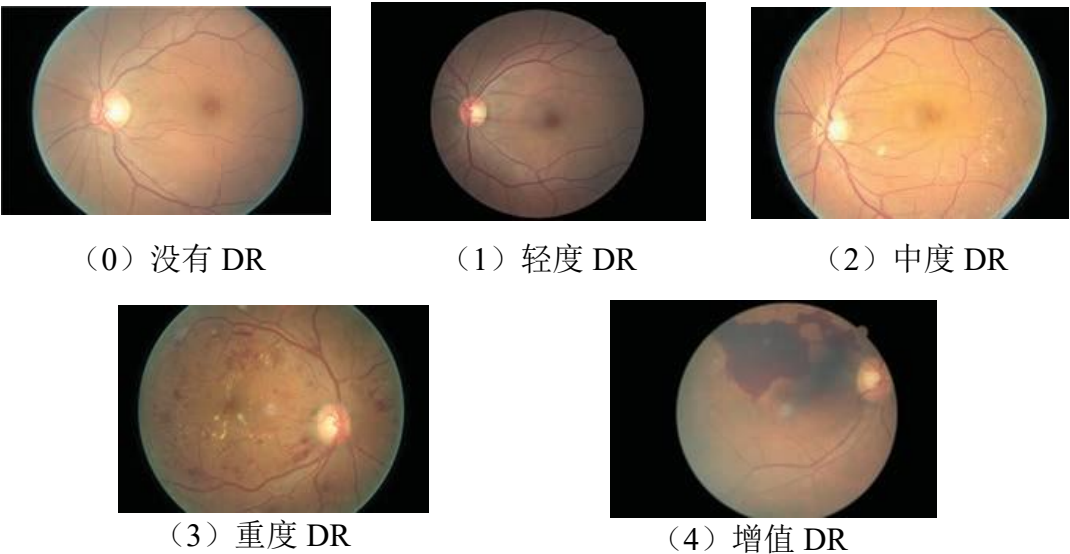


图 2 常见的糖网病患病症状图

Fig.2 Most common lesions caused by diabetic retinopathy

患病程度具体被分为以下 5 类<sup>[3]</sup>：



（3）重度 DR

（4）增值 DR

图 3 糖网病患不同时期的分阶段眼底图像

Fig.3 Stages of diabetic retinopathy (DR) with increasing severity

柏积木 CEO: “对眼底图像的筛选问题一直是人们想要去提升和优化的问题,传统研究的识别方法是二维波斯滤波识别血管、Canny 算子识别视盘、拓扑学找血管,其中应用的最典型的是第三个拓扑学的方法,该方法主要是建立在从形态学去找出血点和微动脉血管瘤。近期对二分类 DR 的方法进行了广泛的研究,比较典型的的就是利用传统 SVM 算法将 DR 就分为患病与不患病两类,最终结果表现虽然特异度很高,但是敏感度很低”。所谓特异度/准确率 (Accuracy rate)<sup>[4]</sup>指:把不患病的人判断为不患病的比例,特异度不高的话,则会造成有很多不患病的人被“误诊”为患病;敏感度/召回率 (Sensitivity)<sup>[4]</sup>指:把患病的人判断为患病的比例,敏感度不高的话,则很多患病的人被“漏诊”为不患病。很显然敏感度很低会导致大量的患病的人被诊断为不患病。理想情况下,要求这两个指标同时较高才能很好的被医院作为诊断决策工具。2000 年到 2012 年左右期间用传统机器学习方法所发表的论文很少同时汇报特异度和敏感度。近 5 年由于计算机图像处理技术的飞速发展,比如 Gardner 等人使用人工神经网络和像素强度值来分别获得 88%特异度和 84.5%敏感度<sup>[5]</sup>,相对而言有了很大的提升。再比如国内的南京航空航天大学的高玮玮、沈建新<sup>[6]</sup>等学者利用 K-means 算法的聚类分析结果对 FCM 进行聚类中心初始化,使得该算法克服了传统 FCM 算法计算复杂度高以及对噪声敏感的缺点。其次,采用两层级联分类结构的 SVM 对候选区域进行分类,也取得了较高的特异度和敏感度。还有来自吉林大学的李丹学者从光学的角度提出的利用光学相干层析技术 (OCT) 作为一种新型的医学成像技术,该技术具有无接触、非入侵、实时活体成像等优点,可以获得清晰的眼底视网膜结构图像,可以监测到视网膜的细微变化,该方法能够有效地分割出眼底 OCT 视网膜图像中能辨别出的所有视网膜层,并且具有较高的准确性和可靠性<sup>[7]</sup>。北京交通大学的丁蓬莉学者提出了一套完整的基于深度神经网络的糖尿病性视网膜图像分类系统,根据视网膜数据集的特殊性提出了一系列图像预处理方式,设计了紧凑的神经网络模型——CompactNet,实验结果表明,CompactNet 网络的微调方法要优于传统的网络训练方法,其分类的准确率可以达到 0.87,与传统直接训练相比高出 0.27<sup>[8]</sup>。在深度学习领域中,通过自我改造深度学习的模型结构来提高分类的准确率以及通过对数据进行放缩、转置、RGB 图像转灰度图以及增加图像对比度等预处理方法来提高模型的精确度。比较有代表性的是电子科技大学的曾杰学者在预处理阶段首先通过主成分分析对眼底图像进行灰度化处理,再用大津阈值法确定视网膜区域;然后进行图像增强来减少光照不均匀的影响和增强对比度等等方法,获得了较高的准确率<sup>[9]</sup>。(采用了 Messidor 数据集,该数据集包含了 89 张糖尿病眼底图像)

正如我们所知,传统的机器学习图像分类方法需要在训练和测试过程中提取所有图像上的特征,这是非常耗时的。比较常用且操作复杂的方法是使用图像处理技术从原始图像中提取血管,微动脉瘤,渗出物和出血,然后将它们送入 SVM 模型进行 5 分类。使用该系统实现了大约 82%左右的灵敏度 (Sensitivity) 86%左右的特异度 (Specificity) 和 86%左右的精度 (Accuracy)<sup>[10]</sup>。

以上这些方法都是在相对较小的数据集上进行的,灵敏度和特异度的下降可能是由于人工提取特征的偏差性和提取不充分所致。以前的 5 分类方法都必须在输入到 SVM 分类器之前从图像中提取特征,并且仅在大约 100 个图像的小测试集上被验证。这些方法总结起来都是建立在数据预处理和模型结构参数调整的基础上进行的,可以发现传统的机器学习算法比如 KNN 和 SVM 算法可以做到不错的准确率,但是前提是要花费大量精力去利用形态学、光学技术和图形处理

技术做特征工程，预先提取出特征再导入模型，而且数据量不能过大，过大容易造成收敛的时间非常漫长，导致模型学习效率低下。最近几年，由于硬件技术的快速发展，伴随着 GPU/CPU 的性能越来越高，在图像识别中使用卷积神经网络（CNN）的实例越来越受到欢迎，与传统方法相比能省去复杂的特征工程，而这一切归功于卷积神经网络卷积层的作用：可以让图片直接输入到网络中，通过内部卷积部分提取出深层特征，大大地简化了人工操作的复杂度。但值得注意的是，利用卷积神经网络模型去做图像识别的重要前提是需要大量的标记好的数据集，每一个实例的 CNN 模型就意味着需要大量该实例的图像数据集，而这个前提在很多领域还无法满足。

### （三）本次研究研究思路与创新之处

#### 1.研究思路

考虑到要训练出一个性能优秀的卷积神经网络模型，需要大量的数据集做支撑，并且从头开始训练模型需要大量的训练数据和时间消耗，故一般在数据集不足的情况下不考虑直接使用卷积神经网络模型。幸运的是，Google 公司提供了一个著名的开源免费图像数据库——ImageNet，该数据库包含了 1000 个类别的百万量级的描绘图像，图像识别模型——Inception\_V3 模型就是基于该数据库训练而来，受到相似领域之间特征相似性思想的启发，本次试验将采用基于源域（Source domain）内 ImageNet 数据集训练出来的 Inception\_V3 模型中卷积部分的网络参数，将目标域（Target domain）内一个含有 324 张标记好的糖网病图像数据集导入该模型，以此来加速提取 324 张图像的深层特征，再将该特征导入到 SVM 多分类模型进行分类，最终得到高准确率分类结果。

整体思路结构示意图

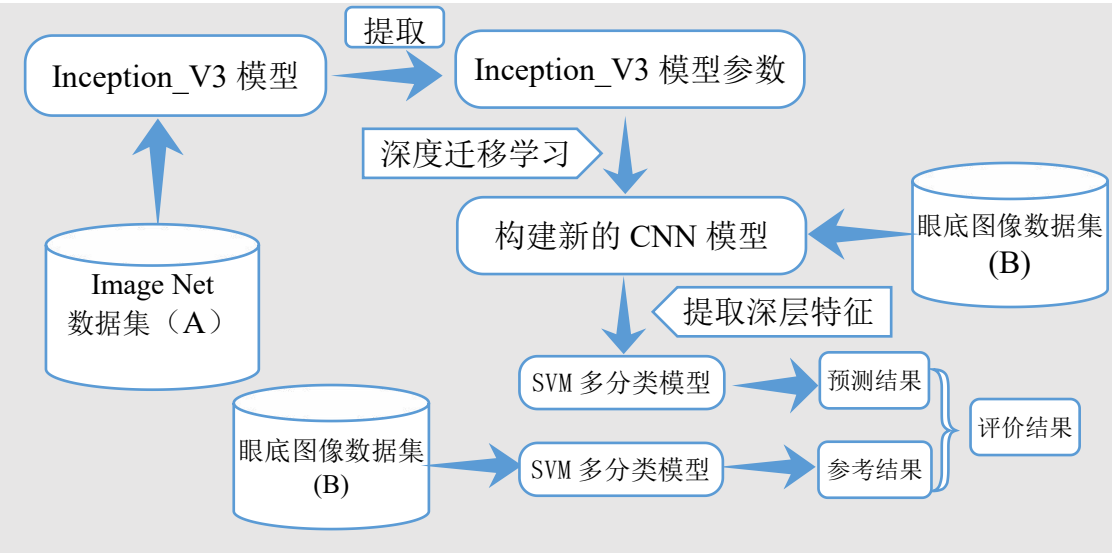


图 4 整体思路结构示意图

Fig.4 The overall structure of the schematic diagram



## 2.创新之处

本次研究采用 CNN+SVM 融合模型的结合算法，来对 324 张眼底图像进行识别分类，充分发挥卷积神经网络的自动提取深层特征功能和深度迁移学习快速收敛的特点，以及少量数据条件下 SVM 模型结构化风险最小的优势，将他们的优势结合利用，达到只需要少量数据集就能做到高特异度、高敏感度的医疗影像分类识别的目标。

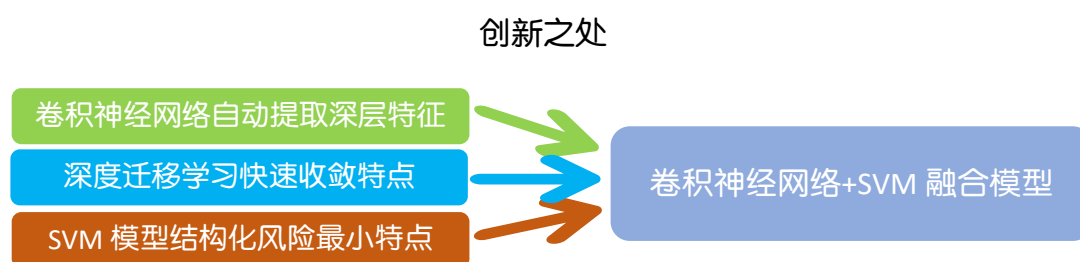


图 5 本次研究的创新之处

Fig.5 The innovation of this study

## 二、方法

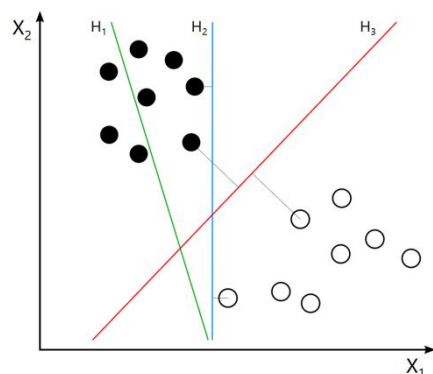
### （一）研究方法的基本原理

#### 1. SVM 模型

SVM 模型(Support Vector Machines)中文全称为支持向量机，是分类与回归分析中分析数据的监督式学习模型与相关的学习算法，支持向量机也被称为：最大间隔分类器<sup>[11]</sup>。

对于支持向量机来说，数据点被视为  $P$  维向量，那么我们要用  $P-1$  维超平面来分开这些点。最佳超平面就是以最大间隔把两类分开的超平面。除了一般的线性分类之外，SVM 还可以使用所谓的核技巧有效地进行非线性分类，将其从低维特征空间映射到高维特征空间去<sup>[11]</sup>。

#### 最大间隔



从图中得知，假设计算机首先找到  $H_1$  直线来作为分割线，发现不能完全分开两类；然后更新法向量（权重  $W$ ），移动  $H_1$  得到了一条可以将数据点完全分开的  $H_2$  直线，但通过计算得知间隔不是最大。依次更新权重  $W$ ，得到最终  $H_3$  直线，以最大的间隔将数据点完全分类。

## 核函数 $\delta$ 从低维映射到高维

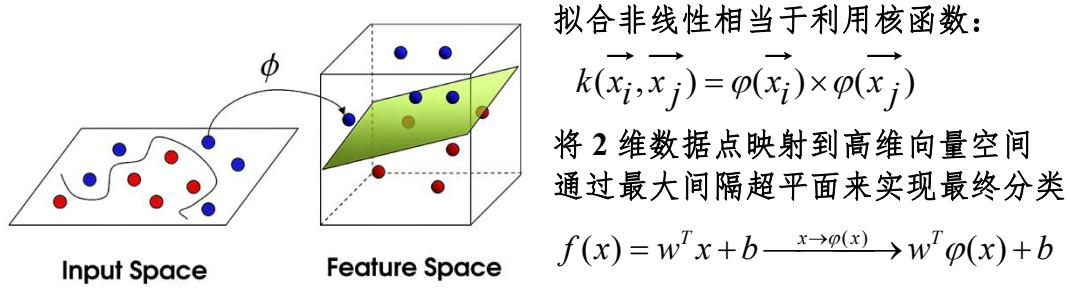


图 6 SVM 分割平面原理

Fig.6 Splitting plane principle of SVM algorithm

本次研究采用 SVM 模型，而不是其他传统机器学习算法，是因为 SVM 算法训练大多数领域的小样本集能够得到比其他算法更优的结果，这得益于他的最大分隔超平面的概念，假设训练数据  $\{(x_i, y_i)\}$ ，根据 SVM 模型的目标函数<sup>[12]</sup>：

$$\min \frac{1}{2} \|W\|^2 + C \sum_i \max(1 - y_i w^T x_i, 0)$$

$\gamma = \frac{2}{\|W\|}$  代表的是两个类别之间的间隔距离， $\max(\frac{2}{\|W\|})$  等价于  $\min(\frac{1}{2} \|W\|^2)$ ，式

中  $\min \frac{1}{2} \|W\|^2$  代表函数的最大间隔， $C \sum_i \max(1 - y_i w^T x_i, 0)$  代表允许错误的大小，

当限制分类错误个数为 C 时， $w^T w \leq C$  等同于 L2 正则化，使得机器学习的 VC 维降低，结构化风险最小，而不是经验风险最小，从而降低了对数据结构和数据分布的要求。直观地，通过与任何类别的距离最近数据点（支持向量）具有最大距离的超平面来实现准确的分离，因为通常数据点离分割线间隔越大的分离器，它的泛化误差越低，结构化风险最小，所以本次研究采用了该模型。

本次研究采用的是多分类 SVM 模型，SVM 算法最初是为二值分类问题设计的，我们可以实现将多分类问题转化为多个二分类问题。常见的方法有“一对多”和“一对一”方法，本次研究采用一对多方法，它是将某个类别的样本归为一类，其他剩余归为另一类，另一类里面又找一类分开，其他又为一类，以此类推，这样就得到多个二分类 SVM 模型，有点类似于决策树结构。

## 2. 卷积神经网络（CNN）

卷积神经网络是深度学习其中的一种模型，深度学习是 Hinton 等人于 2006 年提出，深度学习网络的学习能力比传统机器学习方法强，但是由于模型过于复杂，训练完成一个复杂地模型可能需要花费数日甚至数周的时间才能达到模型收敛，直到 2013 年 Lecun 等人提出的卷积神经网络，它是第一个真正多层结构学习算法，利用局部特征相关关系来减少参数数目以提高训练性能<sup>[13]</sup>。卷积神经网络（CNN）可以被认为是深度学习里面用来图像识别的“尖刀利器”，它包含三个部分，第一部分是输入部分，对于卷积神经网络来说它并不需要像支持向量机或者 BP 神经网络等传统机器学习方法一样：输入之前要进行预先提取代表性特征，在本次试验中，若直接采用 SVM 模型的话，可能要提取出病灶的位置、

形状和颜色等特征作为输入向量，但对于卷积神经网络来说，可以直接把原图像转化成多维数组直接输入到模型，无需手工提取特征；第二部分是卷积部分，它的作用就是利用分类结果的误差反馈自动识别图像特征，然后隐式提取出深层特征；第三部分输出部分，输出部分类似于 BP 神经网络，输入就是卷积部分输出的深层特征，输出的就是最终模型分类的结果。从结构上来看，卷积神经网络（CNN）模型具体包括输入层、卷积层、采样层、全连接层、输出层，其中采样层也叫池化层（Pooling）。结构如下图所示：

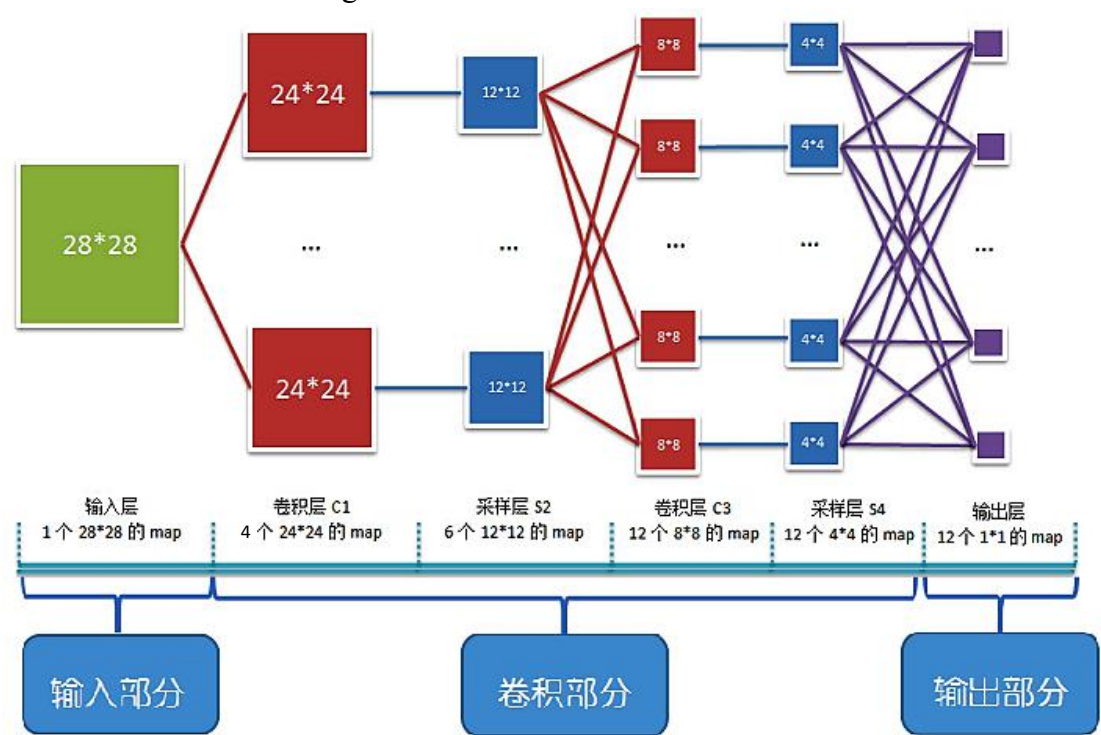


图 7 卷积神经网络的结构

Fig.7 The structure of convolution neural network

顾名思义，卷积部分是该模型的核心，该部分用来提取图片的深层特征，如下图 8 所示，假设输入的是 28\*28 的图像，若采用全连接神经网络的方法，则每个神经元就需要 28\*28 个权重参数，如果 28 个神经元就有 28\*28\*28\*28 的参数（不包括偏移项），可以看到假若图像更大的话，参数数量还会进一步提高，所以神经网络要想进行图像识别，必须减少参数数量。如果由多个卷积核共同来识别图像的话，有意思的事情发生了，我们把卷积层就相当于一排手电筒照在原始图像上，其中卷积核（神经元）就是一个手电筒，4 不同颜色的光代表 4 种不同的卷积核识别不同的特征。卷积神经网络用了两种方法来显著减少训练参数的数量，第一：局部感受野。根据生物学的理论研究可知，一般人对外界的观察是有局部到全局，而图像的空间联系也是局部的像素分布具有联系性，而距离较远的像素联系较弱。因而每一个神经元没有必要对整个图像进行感知，只需要对局部进行感知，然后在后期进行综合得到全局的图像特征，故一个神经元与输入层（28\*28）全连接的方式换成 24\*24 个神经元与输入层的局部区域像素（5\*5）相连，则参数数量有由  $28 \times 28 \times 28 \times 28 \approx 64w$  缩减为  $24 \times 24 \times 5 \times 5 \approx 1w$ ，具体如下图 8 所示：



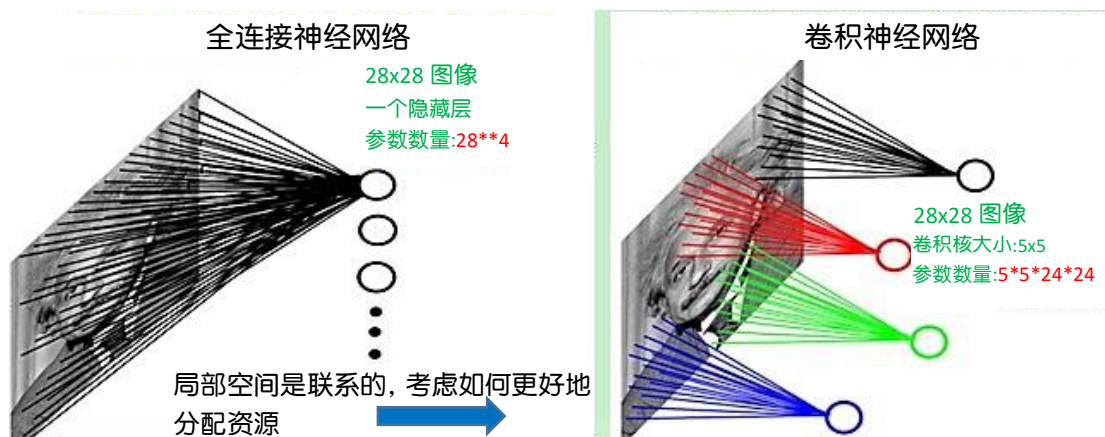


图 8 卷积核的工作原理

Fig.8 The working principle of convolution kernel

第二，参数共享机制。假设一个卷积核向图像射出  $5 \times 5$  面积（感受野）的光。每个卷积核的移动间隔（Stride）是 1 的话，则要想全部罩住一副  $28 \times 28$  的图像，则根据以下公式可以计算出卷积之后的特征图矩阵的大小：

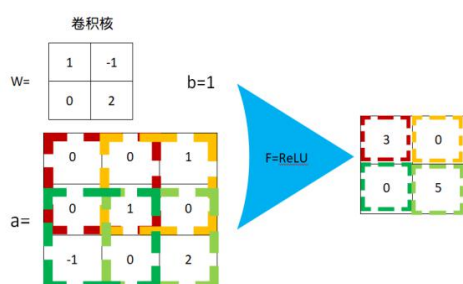
$$feature\_map_{length} = (input_{length} - filter_{length} + 1) / stride_{length}$$

$$feature\_map_{width} = (input_{width} - filter_{width} + 1) / stride_{width}$$

其中  $feature\_map$  就是特征图， $input$  输入图像大小， $filter$  就是卷积核。

通过一个卷积核作用则得到  $\{(28-5+1)/1\} \times \{(28-5+1)/1\} = 24 \times 24$  的特征图，如下图 9 左子图所示。由于参数共享机制，卷积层中的每个卷积核内的权重参数保持共享，所以整个卷积层的可训练参数数量就仅仅为卷积核大小 \* 卷积核个数，由此可见，参数数量进一步由  $24 \times 24 \times 5 \times 5 \approx 1w$  缩减为  $5 \times 5 \times 4 = 120$ ，实际上每个卷积核学习一种特征，对于高清图像来说，比如 RGB 图像，往往需要至少三个卷积核来把图像的颜色特征识别出来。考虑到生成的特征图的可能过大，造成模型进行最终分类之前输入特征的维数过多，容易造成过拟合现象，池化作用是为了进一步压缩特征图大小，其理论基础是相邻图像的特征具有连续性，可以显著地减少参数数量，避免模型出现过拟合现象。如下图 9 右子图所示，例如红色框里面取最大值作为代表特征，假设池化核的大小为  $2 \times 2$ ，Stride 也为 2（红色部分），池化之后的特征图就变成了  $\frac{24 \times 24}{2 \times 2} = 12 \times 12$ ！值得注意的是：池化区域是连续且不重合的，所以在这里 Stride 只能等于 2。综上所述，卷积神经网络解决了参数爆炸的现象，支持高维复杂数据的输入。

### 卷积作用



### 池化作用

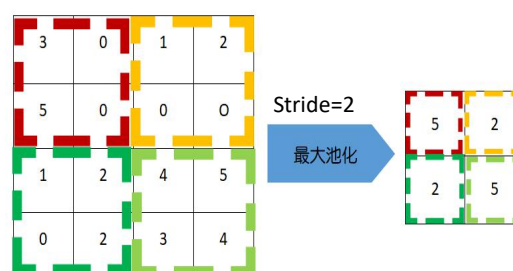


图 9 卷积作用和池化作用原理图

Fig.9 Convolution and pooling of the schematic diagram

为了增强模型的表达能力，通常在每个神经元中加入激活函数，常用的激活函数有 Sigmoid、Tanh、Relu，图 9 中用到的激活函数是 Relu：该函数工作原理如下图所示<sup>[14]</sup>：

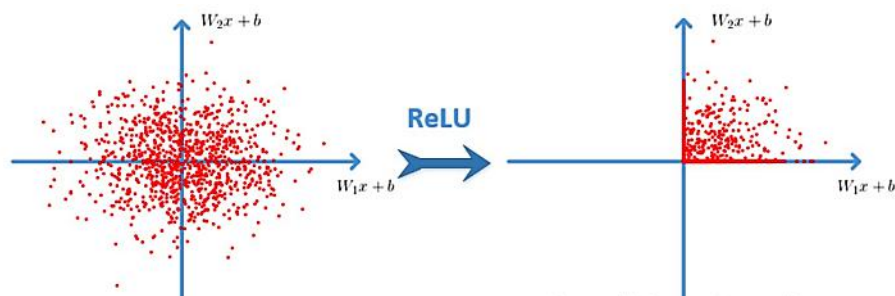


图 10 激活函数的原理图

Fig.10 Activate function of the schematic diagram

假设某一神经元的输入为  $Y = \sum_{i=1}^n w_i x + b$ ，则根据函数

$$\text{ReLU}(Y) = \begin{cases} Y, & w_i x + b \geq 0, i=1,2,3..n \\ 0, & \text{else} \end{cases}$$

故通过卷积作用得到的特征图的左上角数值为 3 是由于：

$$\text{relu}(Y) = \text{relu}(1 \times 0 + (-1) \times 0 + 0 \times 0 + 1 \times 2 + 1) = \text{relu}(3) = 3, \text{其他依次类推};$$

卷积部分提取完特征之后再和全连接层合并，接下来步骤和人工神经网络一相似，也是利用误差后向传播的方法来求最优化的解。值得强调的是：卷积神经网络解决了原先人们担心网络参数太多而导致梯度发散的问题，因为权重共享，所以对于每一个神经元或者卷积层来说，下降梯度会对同一个权值进行修正，积少成多，所以在一定程度也解决了梯度发散的问题<sup>[15]</sup>。

### 3. Inception\_V3 模型

Inception\_V3 的网络结构如下图 11 所示：

| 类型   | kernel 尺寸/步长（或注释） | 输入尺寸           |
|------|-------------------|----------------|
| 1 卷积 | 3*3 / 2           | 299 * 299 * 3  |
| 2 卷积 | 3*3 / 1           | 149 * 149 * 32 |
| 3 卷积 | 3*3 / 1           | 147 * 147 * 32 |
| 4 池化 | 3*3 / 2           | 147 * 147 * 64 |
| 5 卷积 | 3*3 / 1           | 73 * 73 * 64   |
| 6 卷积 | 3*3 / 2           | 71 * 71 * 80   |
| 7 卷积 | 3*3 / 1           | 35 * 35 * 192  |

|                 |                      |                 |
|-----------------|----------------------|-----------------|
| 8Inception 模块组  | 3 个 Inception Module | $35 * 35 * 288$ |
| 9Inception 模块组  | 3 个 Inception Module | $17 * 17 * 768$ |
| 10Inception 模块组 | 3 个 Inception Module | $8 * 8 * 1280$  |
| 11 池化           | $8 * 8$              | $8 * 8 * 2048$  |
| 线性              | logits               | $1 * 1 * 2048$  |
| softmax         | 分类输出                 | $1 * 1 * 1000$  |

图 11 Inception\_V3 的网络结构示意图<sup>[16]</sup>

Fig.11 The diagram of the Inception\_V3 network structure

如图所示 Inception\_V3 模型中设计 CNN 的思想和诀窍：

(1) 将图像输入分开成多个卷积核共同处理的方法很奏效，它可以带来明显减少地参数量，以减轻过拟合现象，达到提升模型的非线性表达能力的效果。

(2) 图中 6 个卷积层贯彻到从输入到输出，其目的是让图片长和宽逐渐减小，输出通道数包括（颜色、透明度）逐渐增加，一个通道代表一种特征。即让空间结构化，将空间信息转化为高阶抽象的特征信息。

(3) 3 个模块组主要是增加模型的拓扑结构的复杂度，分解成多个结构使之能够并行计算，从而在不损失信息的前提下减少了计算量，提高了模型的性能。

#### 4. 迁移学习

许多机器学习和数据挖掘算法的一个主要假设是训练和测试数据必须在相同的特征空间中并具有相同的分布函数。然而，在许多现实世界的应用中，这种假设可能不成立。例如，有时在 A 领域中有一个分类任务 m，该领域内具有大量的训练数据，而在领域 B 也有另有一个分类任务 n，但只有少量的训练数据，后者的数据可能处于不同的特征空间或跟随不同的数据分布。设想如果模型训练得到的特征参数能够转移成功，这将加快模型收敛的速度，不用初始化从头开始训练模型，大大提高学习的效率，这就是应用迁移学习的出发点。形象的描述迁移学习算法，就是将某一领域一个任务上训练好的模型通过调整使其适用于另一领域一个新的任务。

结合目前人工智能发展的趋势，思考为什么需要迁移学习？总结就三点：

- **数据角度**——没有足够数据，如何进行机器学习？

- 收集数据很耗时

- 对数据打标签很难

- 用数据训练模型很繁琐

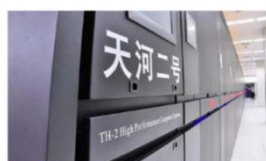


图 12 从数据角度分析进行迁移学习的原因  
Fig.12 Analysis of the reasons for transfer learning from the perspective of data

- **模型角度**——云—端融合模型被广泛应用
  - 通常需要对设备、环境、用户做具体适配
  - 个性化模型适配很复杂
  - 不同的用户隐私处理方式

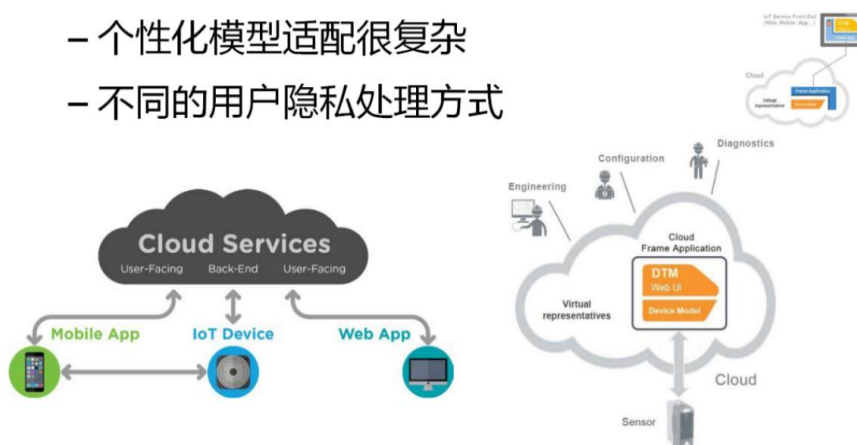


图 13 从模型角度分析进行迁移学习的原因  
Fig.13 Analysis of the reasons for transfer learning from the perspective of model

- **应用角度**——机器学习应用中的冷启动问题
  - 推荐系统：没有初始用户数据，无法精准推荐



图 14 从应用角度分析进行迁移学习的原因  
Fig.14 Analysis of the reasons for transfer learning from the perspective of the application

迁移学习包括两大分类：多源迁移学习和深度迁移学习。其中在图像识别领域深度迁移学习被广泛应用于卷积神经网络（CNN），为高清图像训练的卷积神经网络表现出一个共同的奇怪现象：在第一层，他们学习类似于 Gabor 滤镜和彩色斑点的特征,这样的第一层特征似乎不是针对于特定数据集或任务，而是对于所有图像都存在的特征<sup>[17]</sup>,所以这些卷积层的权重参数自然适用于许多数据集和任务。随着卷积层数的不断增加，自然而然提取到的特征就更细化，更服从特

定种类的图像了。基于卷积神经网络的迁移学习方法的理论基础是相似领域之间的数据具有某些不可变特征成分，或者说是基础特征，如猫和狗都具有四只腿和两只眼睛，可以通过使用训练好猫识别的模型来作为识别狗的基础模型。可以大大地减少对狗数据集数量的依赖。根据论文（DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition）中的结论<sup>[18]</sup>，可以重复利用之前训练好的卷积层的结构参数，只需要替换掉全连接层。同时可以参考 SHL-MDNN [Huang, ICASSP-13][J]<sup>[19]</sup>，该方法描述的是基于卷积神经网络的迁移学习方法进行图像识别。

迁移学习中的符号及概念的定义<sup>[20]</sup>

#### ①符号定义——领域（domain）和任务（task）定义

领域由两个部分组成：特征空间（feature space） $X$  和特征空间的边缘分布  $P(x)$ ，其中， $x = \{x_1, x_2, \dots, x_n\}$  属于  $X$ 。如果两个领域不同，它们的特征空间或边缘概率分布不同。领域表示成  $D = \{X, P(x)\}$ 。任务组成：给定一个领域  $D = \{X, P(x)\}$  的情况下，一个任务也包含两个部分：标签空间  $Y$  和一个目标预测函数  $f(\cdot)$ 。一个任务表示为： $T = \{Y, f(\cdot)\}$ 。目标预测函数不能被直接观测，但可以通过训练样本学习得到。从概率论角度来看，目标预测函数  $f(\cdot)$  可以表示为  $P(Y|X)$ 。任务表示成  $T = \{Y, P(Y|X)\}$

一般情况下，只考虑只存在一个源域  $D_s$  和一个目标域  $D_t$  的情况。其中源领域  $D_s = \{(x_{s1}, y_{s1}), (x_{s2}, y_{s2}), \dots, (x_{sn}, y_{sn})\}$ ， $x_{si}$  属于  $x_s$ ，表示源领域的观测样本， $y_{si}$  属于  $y_s$ ，表示源领域观测样本  $x_{si}$  对应的标签。目标领域  $D_t = \{(x_{t1}, y_{t1}), (x_{t2}, y_{t2}), \dots, (x_{tn}, y_{tn})\}$ ， $x_{ti}$  属于  $x_t$ ，表示目标领域观测样本， $y_{ti}$  属于  $y_t$ ，表示目标领域  $x_{ti}$  对应的输出。通常情况下，源领域观测样本数目  $n_s$  与目标领域观测样本数目  $n_t$  存在如下关系： $1 \leq n_t \ll n_s$ 。

#### ②迁移学习定义

基于以上的符号定义，给出正式的迁移学习的定义：在给定源领域  $D_s$  和源领域学习任务  $T_s$ 、目标领域  $D_t$  和目标领域任务  $T_t$  的情况，且  $D_s$  不等于  $D_t$  或  $T_s$  不等于  $T_t$ ，情况下；迁移学习使用源领域  $D_s$  和  $T_s$  中的知识来提升或优化目标领域  $D_t$  中目标预测函数  $f_t(\cdot)$  的学习效果。通过以上的定义可以发现：

（一）领域  $D = (X, P(x))$ ，当源和目标领域  $D$  不同时，存在两种情况：（1） $X_s$  不等于  $X_t$ ，源领域和目标领域的特征空间不同；（2） $P(x_s)$  不等于  $P(x_t)$ ，即源空间和目标空间的特征空间相同，但观测样本  $X$  的边缘分布不同。

（二）任务  $T = \{Y, P(Y|X)\}$ ，当源和目标领域  $T$  不同时，存在两种情况：（1） $Y_s$  不等于  $Y_t$ ，源领域的标签空间与目标领域的标签空间不同；（2） $P(Y_s|X_s)$  不等于  $P(Y_t|X_t)$ ，即源领域和目标领域的条件概率分布不同。

| 总结   | 传统机器学习     | 迁移学习          |
|------|------------|---------------|
| 数据分布 | 训练和测试数据同分布 | 训练和测试数据不需要同分布 |
| 建模标签 | 足够的数据标注    | 不需要足够的数据标注    |
| 建模   | 每个任务分别建模   | 模型可以反复使用      |

表 1 传统机器学习与迁移学习的区别

Table.1 The difference between traditional machine learning and transfer learning



基于以上理论和方法，我们将源域  $D_s$  中的 ImageNet 数据集输入到模型——Inception\_V3 中，获取到模型参数文件(classify\_image\_graph\_def.pb)，再将目标域  $D_t$  中 2250 张已经标注好糖网病影像数据集作为输入，再次调用该模型结构参数文件，提取出深层特征，然后输入到 SVM 多分类模型得到最终分类结果，并将其与预处理后直接导入 SVM 模型的方法结果相比较，以此来验证迁移学习的可行性和迅速收敛的优势。

## （二）数据的来源

本次研究的数据来源分两部分：第一部分数据为了训练 Inception\_v3 模型而采用了 ImageNet 数据集，该数据集包含 120 万标注图片，被分为 1000 个类别的图像；第二部分数据来自山东省泰安市某医院所提供的 324 位真实患者的眼底图像数据,像素为 2592 x 1728。其中没有 DR 有 246 例，轻度 DR 有 1 例，中度 DR 有 30 例，重度 DR 有 3 例，增值 DR 有 44 例。

## （三）数据的预处理

由于患者的年龄段不一样、医生拍摄个人习惯和角度把握不恰当等问题，影响到图像中的像素强度值以及曝光度，与此同时还包括大面积的个人隐私信息分布在眼底图像周围，如果不进行处理将会对最终结果产生影响，为尽量消除这一影响，使用 PIL (<http://www.pythonware.com/products/pil/>) 专业的图像处理包来分别对图像去黑边化、normal\_size 化、灰值图像化处理。

### 1. 去除图像周围的黑边和个人隐私信息

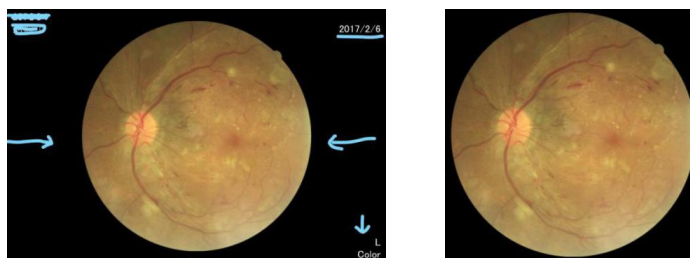


图 15 去除图像周围的黑边和个人隐私的效果图

Fig.15 Removed the black border and personal privacy around the image

### 2. 将图像大小都化为 512\*512

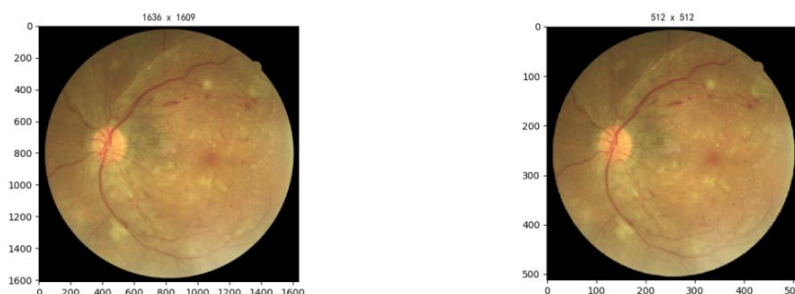


图 16 将图像大小化为 512\*512 大小的效果图  
Fig.16 Resized the image to 512 \* 512 size

### 3. 将 RGB 图像转化为灰值图像

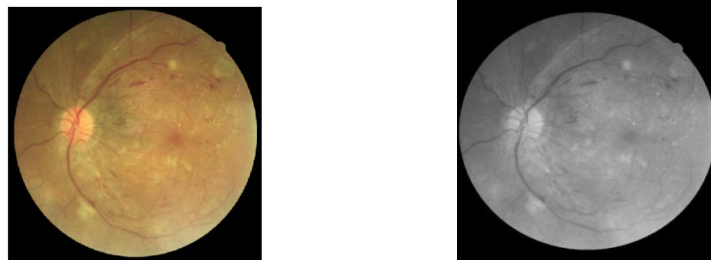


图 17 将 RGB 图像转为灰值化图像的效果图  
Fig.17 Illustration of converted an RGB image to a grayscale image

## （四）数据扩增技术

按数据的预处理那章节所述，324 例中没有 DR 的有 246 例，轻度 DR 有 1 例，中度 DR 有 30 例，重度 DR 有 3 例，增值 DR 有 44 例。很明显患者患病的类别数量及其不均衡，轻度 DR 只有 1 例，而没有 DR 则高达 246 例。若是直接将它导入模型的话，则会造成模型过度识别为没有 DR，而不易学习到轻度 DR 特征的情况，所以考虑在不增加数据源的条件下，在源数据基础上通过图像处理技术来增加数据集的规模，再与源数据集合并导入模型，此方法称为源数据扩增处理。

本次研究通过对源数据进行图像直方图化、图像旋转、以及像素点位加强的手段来增加数据量。

### 1. 图像直方图化

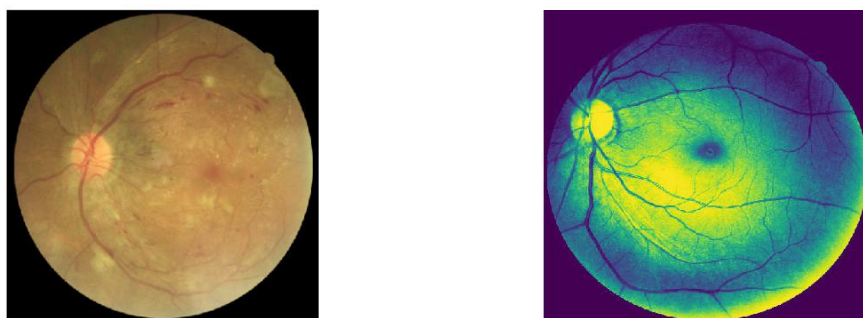


图 18 图像直方图化的效果图  
Fig.18 Illustration of histogram equalization for the image

图像直方图化不仅可以衍生出同等数量的数据集，而且从特征识别的角度来说，图像直方图化会加大图像的对比度，有利于特征的提取，从而提高识别准确率。

2. 图像旋转



图 19 图像旋转之后的效果图  
Fig.19 Illustration of image rotation

3. 每个像素值增加到 2 倍



图 20 每个像素值增加到 2 倍后的结果图  
Fig.20 Illustration of Each pixel value increased to 2 times the original

总的来说，包括之前的灰度图像，现在源数据量已经扩大到 5 倍，数量达到了 1620。整个数据集合在一起如下图所示：

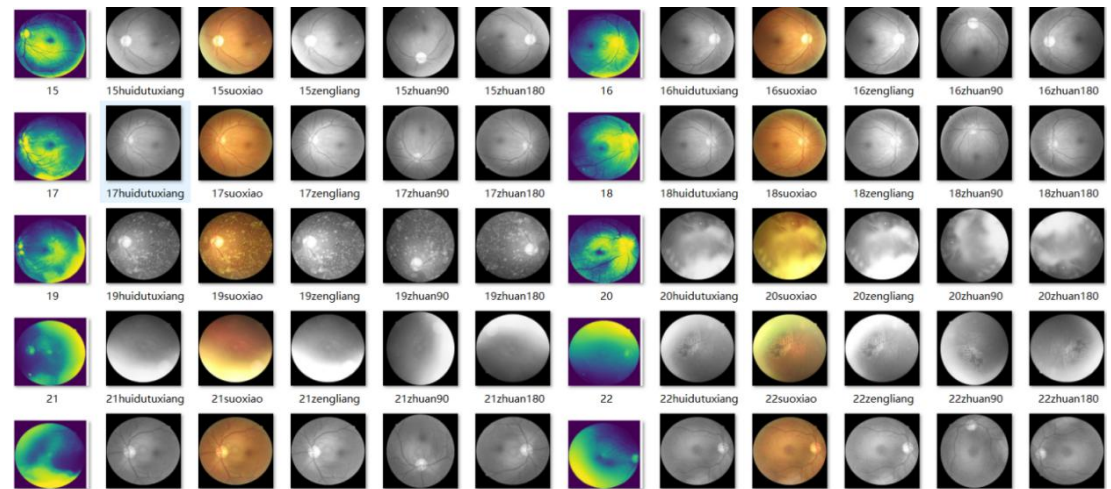


图 21 将原数据扩大 5 倍后的数据示例图  
Fig.21 Illustration of the original data set to expand 5 times

## （五）模型的构建

### 1. SVM 模型

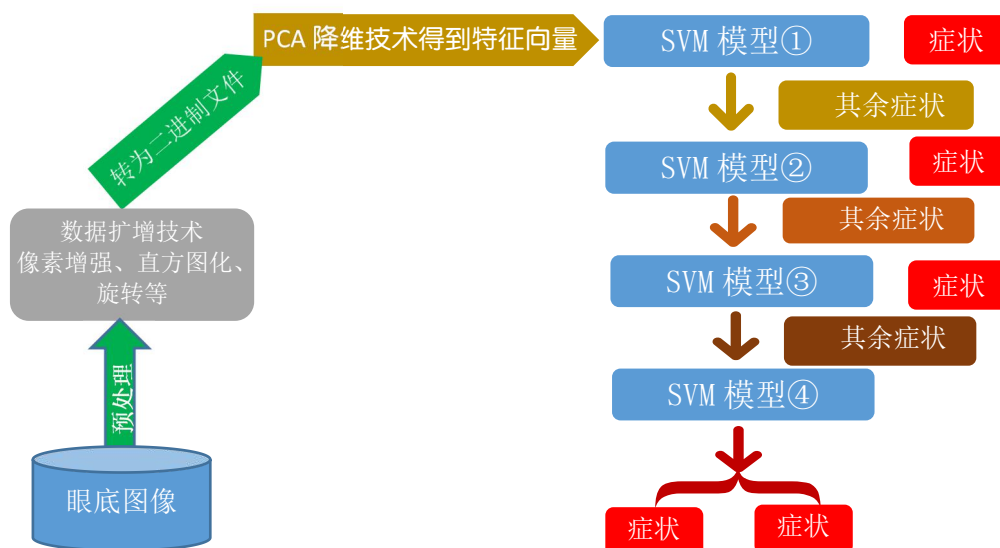


图 22 SVM 模型识别过程图

Fig.22 The process diagram of SVM model identification

#### 预处理：

①图像一律剪切成 512\*512 大小，图像已去除个人隐私信息

②RGB 图像进行灰度化处理，减少通道数，降低后续计算量

#### 数据扩增技术：

由于数据量过少，且每种类型样本量严重不均衡。扩增情况如下：

经过图像旋转、图像直方图化、图像像素加强等方法进一步扩增到 2250 张图片，其中没有 DR 的有 1230 例，轻微 DR 的有 300 例，中度 DR 有 230 例，重度 DR 的有 270 例，增值 DR 的有 220 例。将整个数据集分成 70%训练集，30%测试集，最后利用随机抽样的方法，把每一类样本数量都变成 1230.

#### PCA 降维技术：

PCA：指的是主成分分析，挑选出特征值最大的几个特征向量来代替之前的特征向量，保证保留的维度也能充分代表图像特征，以此来降低每一样本的维度，减少计算量，能够有效防止出现过拟合现象。本次利用 sklearn 库中的 decomposition 包中 PCA 模块将  $512 \times 512 = 262144$  维特征向量降到 150 维。参数优化方法选的是 SVD，其中参数 `svd_solver`：即指定奇异值分解 SVD 的方法，有 4 个可以选择的值：{ 'auto', 'full', 'arpack', 'randomized' }。randomized 一般适用于数据量大，样本的维数多同时主成分比例又较低的 PCA 降维，它使用了一些加快 SVD 的随机算法。故本次选择的是 randomized 参数。

#### SVM 多分类模型：

方法：“一对多”方法（OVR），将 sklearn.multiclass 中 OneVsRestClassifier 包以及 SVM.SVC 相结合的方法来构建多分类模型。

核函数：因为数据样本点不能完全被线性分开，且维数较高，所以采用的核函数是高斯核函数（RBF），拟合非线性数据，且能够减少参数可以有效防止过拟合。

**惩罚项参数 C:** 惩罚项参数 C 越大，相当于惩罚松弛变量，希望松弛变量接近 0，即对误分类的惩罚增大，趋向于对训练集全分对的情况，这样对训练模型的分类结果准确率要求很高，但模型泛化能力弱。C 值小，对误分类的惩罚减小，允许容错，将他们当成噪声点，泛化能力较强。本次取默认值 C=1，强调特异度和敏感度都要达到一定高的水平。

## 2. 卷积神经网络+SVM 融合模型

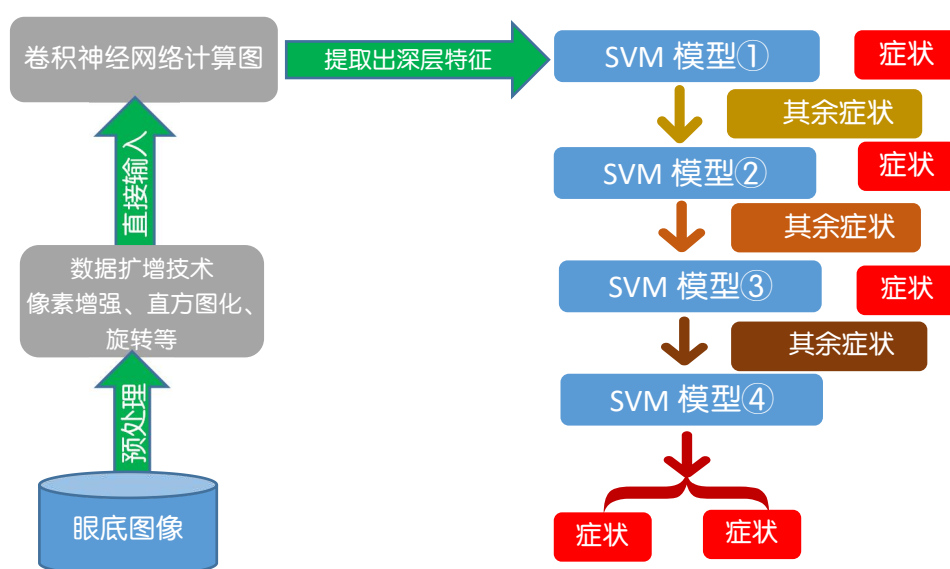


图 23 卷积神经网络与 SVM 融合模型的识别过程

Fig.23 Recognition process of convolution neural network and SVM fusion model

**预处理和数据扩增与 SVM 模型处理一样**

**卷积神经网络:** 从 Inception\_V3 模型提取训练好的计算图文件 (.pb), 该文件保留了 ImageNet 数据集的特征向量的所有参数, 加载该计算图文件, 直接输入 2250 张眼底图像, 利用 feature\_extraction 模块提取出深层特征作为 SVM 模型的输入层。

**数据划分:** 70%训练集, 30%测试集。

为保证两模型最终的训练和预测效果, 所以 SVM 模型部分的参数设置与前一模型保持一致。

## 三、结果

### (一) 模型结果可视化

分类报告

SVM 模型

CNN+SVM 模型



| 分类报告      | precision   | recall      | F1-score    | support    |
|-----------|-------------|-------------|-------------|------------|
| 没有 DR     | 0.71        | 0.72        | 0.71        | 369        |
| 轻微 DR     | 0.63        | 0.70        | 0.66        | 90         |
| 中度 DR     | 0.10        | 0.07        | 0.08        | 69         |
| 重度 DR     | 0.74        | 0.79        | 0.77        | 81         |
| 增值 DR     | 0.10        | 0.09        | 0.09        | 66         |
| avg/total | <b>0.58</b> | <b>0.60</b> | <b>0.59</b> | <b>675</b> |

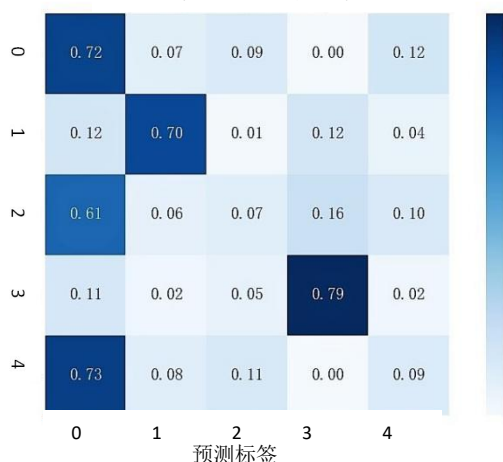
| 分类报告      | precision   | recall      | F1-score    | support     |
|-----------|-------------|-------------|-------------|-------------|
| 没有 DR     | 0.81        | 0.75        | 0.78        | 369         |
| 轻微 DR     | 0.96        | 0.95        | 0.95        | 369         |
| 中度 DR     | 0.86        | 0.89        | 0.88        | 369         |
| 重度 DR     | 0.97        | 0.99        | 0.98        | 369         |
| 增值 DR     | 0.86        | 0.88        | 0.87        | 369         |
| avg/total | <b>0.89</b> | <b>0.89</b> | <b>0.89</b> | <b>1845</b> |

从两幅分类报告对比看出，对于 SVM 模型来说 CNN+SVM 模型每项数据均至少优于 50%。

|                  |   |     |    |    |    |    |
|------------------|---|-----|----|----|----|----|
| 真<br>实<br>标<br>签 | 0 | 124 | 26 | 35 | 0  | 44 |
|                  | 1 | 11  | 63 | 1  | 11 | 4  |
|                  | 2 | 42  | 4  | 5  | 11 | 7  |
|                  | 3 | 9   | 2  | 4  | 64 | 2  |
|                  | 4 | 48  | 5  | 7  | 0  | 6  |
|                  |   | 0   | 1  | 2  | 3  | 4  |
| 预测标签             |   |     |    |    |    |    |

|                  |   |     |     |     |     |     |
|------------------|---|-----|-----|-----|-----|-----|
| 真<br>实<br>标<br>签 | 0 | 276 | 9   | 39  | 2   | 43  |
|                  | 1 | 12  | 350 | 3   | 3   | 1   |
|                  | 2 | 26  | 1   | 330 | 3   | 9   |
|                  | 3 | 1   | 2   | 0   | 364 | 2   |
|                  | 4 | 26  | 3   | 11  | 4   | 325 |
|                  |   | 0   | 1   | 2   | 3   | 4   |
| 预测标签             |   |     |     |     |     |     |

SVM 多分类模型的混淆矩阵 seaborn 图



SVM+CNN 多分类模型的混淆矩阵 seaborn 图

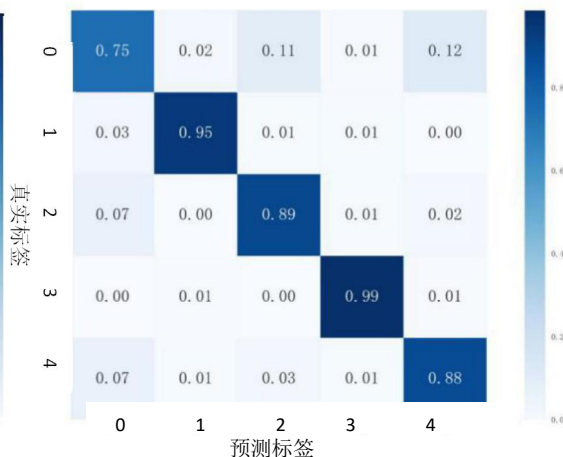


表 2 SVM 和 CNN+SVM 融合模型识别分类结果分析

Table.2 Comparison and analysis of SVM and CNN + SVM fusion model recognition and classification results

混淆矩阵 seaborn 图和混淆矩阵不同的是：主对角线上的百分比就是敏感度（召回率），从混淆矩阵 seaborn 图可以很清楚的看出两个模型分类结果的敏感度的大小。左边 SVM 模型在鉴定没有 DR（0 级）、轻微 DR（1 级）、重度 DR（3 级）方面表现不错，它正确地识别了 72%、70%、79% 这些类中标记的图像。右边 CNN+SVM 模型表现的更好，基本上每一类的敏感度都很高。把患病的级别都分的很好，把患病的大多数鉴定为患病，把浅色部分加起来只有 17% 的比例把患病的人判断为不患病。对于 SVM 模型来说，该模型在区分糖尿病性视网膜病变（1-4 级）和没有疾病迹象（0 级）方面表现不太好。比如说 61% 的中度 DR 患者被鉴定为不患病，高达 73% 的 4 级患者被鉴定为不患病，为了使模型在临床环境中有用，不仅应该正确识别疾病的种类，而且应尽可能减少假阴性的数量（患者患有 DR，但模型未能鉴别的情况），若敏感度不高的话，患者可能不会被送去进一步的调查和治疗，从而延误了病情，不应被用于临床医疗环境。

## （二）模型预测性能比较

| 模型       | SVM 模型     | CNN+SVM 模型 |
|----------|------------|------------|
| 训练模型花费时间 | 1993.36（秒） | 1518.46（秒） |
| Kappa 值  | 0.20       | 0.79       |
| 特异度      | 0.58       | 0.89       |
| 敏感度      | 0.60       | 0.89       |

由于深度迁移学习方法的使用，无论是训练时间，还是 kappa 值，或者特异度和敏感度都远远优于单纯的 SVM 模型。

表 3 中，判断模型预测性能首先参考标准是 Kappa 值大小，Kappa 值，描述的是样本预测分类与真实分类的匹配程度，一般 Kappa 值属于-1 到 1 之间，但通常结果落在 0 到 1 之间，用字母 k 来表示： $k = \frac{p_0 - p_e}{1 - p_e}$ ， $p_0$  为总体分类精度，以

CNN+SVM 为例：

$$\begin{aligned}
 p_0 &= \frac{276 + 350 + 330 + 364 + 325}{369 \times 5} = 0.892 \\
 p_e &= \frac{369 \times (276 + 12 + 26 + 1 + 26) + \dots}{369 \times 5 \times 369 \times 5} = 0.486 \\
 k &= \frac{0.892 - 0.486}{1 - 0.486} = 0.79
 \end{aligned}$$

表 3 SVM 和 CNN+SVM 模型预测性能比较

Fig.3 Comparison of predictive performance between SVM and CNN + SVM fusion model

## 四、启示和建议

### （一）启示

当 A 领域中的任务 m 出现数据集不够的情况时，考虑利用迁移学习方法，使用相似 B 领域中的拥有海量数据的任务 n 预先训练图像识别模型，然后通过改造最后的全连接层来适应任务 m，全连接层考虑用 SVM 模型或者 BP 神经网络和随机森林等等这些善于处理低维、少量数据集的模型来做最后的分类，这样就能实现从 B 领域中的任务 n 训练出来的基础图像的相关特征信息迁移到 A 领域中的任务 m 上，由此可见深度迁移学习可以大大增强模型的泛化能力。迁移学习的好处就是只需要花费极少的时间——“站在巨人的肩膀上学习”也能够得到高准确率图像识别模型。基于此研究的最终结果得出：通过在相似的领域之间使用深度迁移学习的方法，可以实现性能优越的图像自动筛选系统的功能，那么同样可以推断出考虑基于 Inception\_V3 模型来自动识别青光眼疾病、红眼病和其他基于 CT 影像诊断的疾病症状，根据相似性原理最终结果与此次试验的结果接近，从而构成一套完整的疾病自动筛选系统，为医生提供辅助性的诊断决策，有效发挥出“第二医生”的作用。

## （二）建议

国家层面应该积极支持医疗影像自动筛选系统的发展，实施严格化、标准化的法律措施来统一制定医疗影像数据的产生、存储、再开发的完整系统流程，以此来带动整个医疗行业向着平衡医疗资源的方向发展；医疗企业层面应该重视医疗影像数据的收集与存贮工作，积极地拥抱大数据技术，与互联网公司展开深度、密切合作，最终实现数据收集到医疗辅助诊断系统的无缝对接；老百姓层面应该给予大数据技术的足够信任度，消除对机器化的偏见，同时积极地做好疾病的早预防、定期检查的工作，为行业不断提供数据源，利用这些最真实的患者数据产生医疗影像自动分类和提供辅助诊断决策系统，最终达到减少误诊率、漏诊率的目的，为全社会的医疗发展做出积极贡献。

## 五、研究不足与展望

### （一）研究不足

①数据集的样本数量分布不均，存在着过多的没有患病的眼底图像，而轻度 DR 只有 1 例，严重 DR 只有 3 例，不可避免地导致模型过分学习了没有 DR 的图像特征，而学到很少的其他类型的特征，所以才会出现 SVM 模型中度 DR 和增值 DR 被模型预测成没有 DR 的比例分别为 61%和 73%，导致 kappa 值过小的情况。若是原始样本数量分布均匀的话，两个模型的 kappa 值还可以做到更好。

②本次研究拿到的是数据集可能由于医生的个人习惯、图片拍摄的方法的差异和设备条件的限制等原因造成数据集的标记的质量有所下降。

③由于本次研究用于运行代码的工具的配置有限（没有使用 GPU），限制了模型不能过于复杂，从某种程度上讲，也损失了一些准确率。

### （二）展望

随着人工智能的迅速发展，即将诞生许多的传统项目与大数据技术所结合的大数据工程。企业开始重视内部的大量数据资源，所以对于数据难获取的某些行业以及欠发达地区来说，迁移学习可以解决一些问题，显然迁移学习会朝着更加细致化的方向发展，三个方向：（1）迁移什么、（2）如何迁移、（3）什么时候迁移。

（1）迁移什么：找到多个领域共同的基础特征，通过源域学习这部分特征，来迅速提高学习的效率；

（2）如何迁移：如何迁移才能获得更多的特征知识，这根具体的数据类型有关，比方说带标记的图像数据倾向于深度神经网络，无标记的数据比如领域不存在或者很难获取带标记数据（比如航空航天、战争死亡率、非洲国家的数据），则考虑（聚类、PCA 降维，t\_SNE 降维）等无监督迁移学习方法来构建迁移学习模型；由于对源数据的依赖更少，更符合实际情况，所以不久的将来无监督迁移学习的发展将会取到巨大进展；

（3）什么时候迁移：首先不能盲目的从一领域迁移到另一不相干领域，迁移和源域与目标域的相似度多大才能迁移，随着以后理论的发展，一定要量化出来具体的临界指标，否则的话会造成知识负转移现象。

## 参考文献

- [1] 马艳红. 我国“糖网病”筛查率不足 10%[N]. 中国医学报, 2016, (3).  
(MA Y H. "Diabetic Retinopathy" screening rate of less than 10% in China[N]. Chinese Journal of Medical Science. 2016, (3))
- [2] Quellec G, Charrière K, Boudi Y, et al. Deep image mining for diabetic retinopathy screening[J]. Medical Image Analysis, 2017, 39: 178-193.
- [3] Pratt H, Coenen F, Broadbent D M, et al. Convolutional neural networks for diabetic retinopathy[J]. Procedia Computer Science, 2016, 90: 200-205.
- [4] 佚名. 全科医生小词典——敏感度与特异度[J]. 中国全科医学, 2013(14):1668-1668.  
(Anon. Chinese General Practice-Sensitivity and specificity[J]. Journal of Modern Medicine and Health, 2013, (14):1668-1668.)
- [5] Gardner G G, Keating D, Williamson T H, et al. Automatic detection of diabetic retinopathy using an artificial neural network: a screening tool[J]. British journal of Ophthalmology, 1996, 80(11): 940-944.
- [6] 高玮玮, 沈建新, 王玉亮, 等. 改进的快速 FCM 及 SVM 实现糖网白色病灶的自动检测[J]. 中国生物医学工程学报, 2013, 32(3): 305-312.  
(GAO W W, SHEN J X, WANG Y L, et al. Automated Detection of Bright Lesions of Diabetic Retinopathy Based on Improved and Fast FCM and SVM[J]. Chinese Journal of Biomedical Engineering, 2013, 32(3): 305-312.)
- [7] 李丹. 基于连续最大流算法眼底 OCT 图像视网膜层分割方法研究[D]. 吉林大学, 2016:29-37.  
(LI D, Study on Retraction Method of OCT Image Retinal Layer Based on Continuous Maximum Flow Algorithm[D], Jilin University, 2016:29-37.)
- [8] 丁蓬莉. 基于深度学习的糖尿病性视网膜图像分析算法研究[D]. 北京交通大学, 2017:42-48.  
(DING P L, Research on Diabetic Retinal Image Analysis Algorithm Based on Deep Learning[D]. Beijing Jiaotong University, 2017:42-48.)
- [9] 曾杰. 视网膜图像的分析与研究[D]. 电子科技大学, 2016.  
(ZENG J, Analysis and research of retinal images[D]. University of Electronic Science and Technology, 2016.)
- [10] Quellec G, Charrière K, Boudi Y, et al. Deep image mining for diabetic retinopathy screening[J]. Medical Image Analysis, 2017, 39: 178-193.
- [11] <https://zh.wikipedia.org/wiki/SVM>

- [12]周志华. 机器学习[M]. 清华大学出版社, 2016:121-129.  
(ZHOU Z H, Machine learning[M]. Tsinghua University Press, 2016:121-129.)
- [13]LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [14]Clevert D A, Unterthiner T, Hochreiter S. Fast and accurate deep network learning by exponential linear units (elus)[J]. arXiv preprint arXiv:1511.07289, 2015.
- [15]Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]//International Conference on Machine Learning. 2015: 448-456.
- [16]Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 2818-2826.
- [17]Yang M, Liu Y, You Z. The Euclidean Embedding Learning based on Convolutional Neural Network for Stereo Matching[J]. Neurocomputing, 2017.
- [18]Donahue J, Jia Y, Vinyals O, et al. Decaf: A deep convolutional activation feature for generic visual recognition[C]//International conference on machine learning. 2014: 647-655.
- [19]Huang J T, Li J, Yu D, et al. Cross-language knowledge transfer using multilingual deep neural network with shared hidden layers[C]//Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on. IEEE, 2013: 7304-7308.
- [20]Pan S J, Yang Q. A survey on transfer learning[J]. IEEE Transactions on knowledge and data engineering, 2010, 22(10): 1345-1359.