

肝脏扫描叶段分类和增强

XXX*
XX 学院
XX 大学
XXX
XXX

XXX*
XX 学院
XX 大学
XXX
XXX

XX*
XX 学院
XX 大学
XXX
XXX

XXX*
XX 学院
XX 大学
XXX
XXX

摘要—本文通过使用 RestNet、GoogLeNet、AlexNet 三种神经网络模型实现了将超声图像按照标注的扫描叶段进行分类的任务，首先在训练集上对三种模型分别进行训练，最终使用训练好的模型在测试集中进行实验，并且将三种网络模型的实验结果进行全面的对比；此外对于数据集图像本文使用相对全局直方图拉伸(RGHS)的方法进行图像增强，经过对比度校正和颜色校正后，该方法可以保留图像细节，提高图像的可视性。

关键字—肝脏扫描叶段分类，图像增强，神经网络模型

I. 引言

随着医学成像技术和计算机技术的不断发展和进步，医学图像分析已成为医学研究、临床疾病诊断和治疗中一个不可或缺的工具和技术手段。近几年来，深度学习特别是深度卷积神经网络已经迅速发展成为医学图像分析的研究热点，它能够从医学图像大数据中自动特区隐含的疾病诊断特征。肝脏分段可以了解病灶的位置，特别是对与肝脏的恶性肿瘤而言可以指导治疗。然而传统的肝脏分段需要人工进行，这不仅耗时耗力同时对于观察人的专业性有着很高的要求，因此利用神经网络模型，自动将超声图像按照标注的扫描叶段进行分类对于疾病诊断与治疗有着巨大的帮助，不仅省时省力而且准确率也比较高。

目前使用神经网络模型进行图像的分类技术发展已经相当成熟。1998 年推出 LeNet[1]，为未来使用卷积神经网络的图像分类研究奠定了基础；2012 年 ImageNet 大规模视觉识别挑战赛中，Alex Krizhevsky 提出了 AlexNet 一种基于 CNN 的解决方案，并将 ImageNet 测试集 top-5 的准确率从 73.8% 大幅提高到 84.7%。他们的方法继承了 LeNet 的多层 CNN 思想，但大大增加了 CNN 的规模；2014 年 VGG[2] 被提出用于大型图像识别的超深度卷积网络，VGG 外形美观，结构简单易懂，但在 ImageNet 2014 年的所有决赛中，它的表现并不是最好的，而是 GoogLeNet，又名 InceptionV1，赢得了最后的奖项；2015 年诞生了一个新的网络，称为 ResNet，或称为残差网络，该网络由 Microsoft Research Asia 的一组中国研究人员提出；由于跨通道卷积通常会引入大量参数，因此 2016 年提出的 Xception[3] 网络选择调查此操作以了解其效果的全貌；2017 年 MobileNet[4] 在资源受限的环境中引领了图像分类的重大进展。在本文中，主要选取了 RestNet、GoogLeNet、AlexNet 三种代表性的网络模型，来将超声图像按照标注的扫描叶段进行分类。

II. 肝脏扫描叶段分类

A. AlexNet

在 2010 年的 ImageNet LSVRC-2010 上，AlexNet [5] 在给包含有 1000 种类别的共 120 万张高分辨率图片的分类任务中，在测试集上的 top-1 和 top-5 错误率为 37.5% 和 17.0%（top-5 错误率：即对一张图像预测 5 个类别，只要有一个和人工标注类别相同就算对，否则算错。同理 top-1 对一张图像只预测 1 个类别），在 ImageNet LSVRC-2012 的比赛中，取得了 top-5 错误率为 15.3% 的成绩。

AlexNet 创新之处在于：

- 成功使用 ReLU 作为 CNN 的激活函数，验证了其效果在较深的网络中超过了 Sigmoid，成功解决了 Sigmoid 在网络较深时的梯度弥散问题。
- 训练时使用 Dropout 随机忽略一部分神经元，以避免模型过拟合，一般在全连接层使用，在预测的时候是不使用 Dropout 的，即 Dropout 为 1。
- 在 CNN 中使用重叠的最大池化(步长小于卷积核)。此前 CNN 中普遍使用平均池化，使用最大池化可以避免平均池化的模糊效果。同时重叠效果可以提升特征的丰富性。
- 提出 LRN 层，对局部神经元的活动创建竞争机制，使得其中响应比较大的值变得相对更大，并抑制其他反馈较小的神经元，增强了模型的泛化能力。

AlexNet 网络结构共有 8 层，前面 5 层是卷积层，后面 3 层是全连接层，最后一个全连接层的输出传递到一个 1000 路的 softmax 层，对应 1000 个类标签的分布。

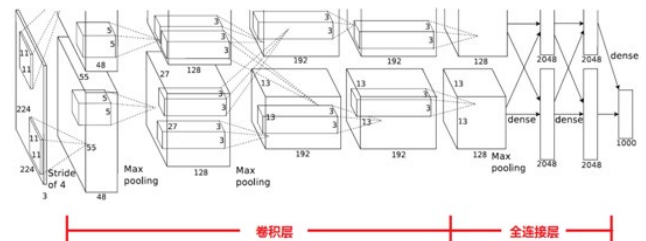


图 1 AlexNet 网络结构

B. GoogleNet

Google 公司在 2014 年提出了一种改进的卷积神经网络 GoogLeNet[6]，获得了 ImageNet 竞赛的冠军。GoogLeNet 与 VGG 等其他卷积神经网络的区别在于使用

了 Inception，将多个卷积操作（主要是 3*3 卷积和 5*5 卷积）拼接在一起组成了一个网络模块。对于同一张图片，不同尺寸的卷积核的表现效果是不一样的，因为他们的感受野不同。Inception 的加入可以使网络在训练的过程中通过调节参数自己去选择使用何种卷积核。不同 Inception 模块的卷积部分有不同的通道数设置，整个 GoogLeNet 是许多 Inception 模块的组合。

Inception 的结构如图 2 所示，结果由两个 1*1 卷积、一个 3*3 卷积、一个 5*5 卷积拼接而成。四个卷积过程中都只是较少了通道数，没有对图片尺寸做出改变。

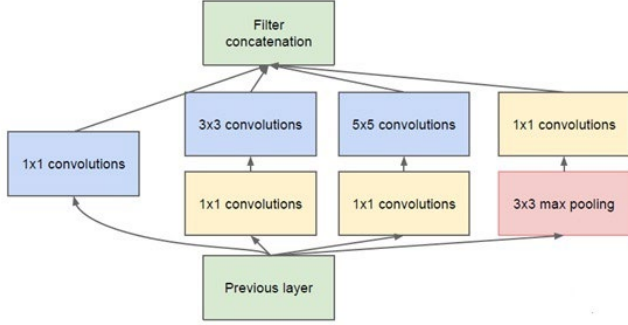


图 2 Inception 结构

标准的 GoogLeNet 的结构如图 3 所示，包含 5 个阶段。输入图片的尺寸为 3*244*244，第 1 阶段得到的图片尺寸为 64*56*56，第 2 阶段得到的图片尺寸为 192*28*28，第 3 阶段得到的图片尺寸为 480*14*14，第 4 阶段得到的图片尺寸为 832*7*7，第 5 阶段得到的图片尺寸为 1024*7*7，经过全局池化得到输出 1024*1*1。

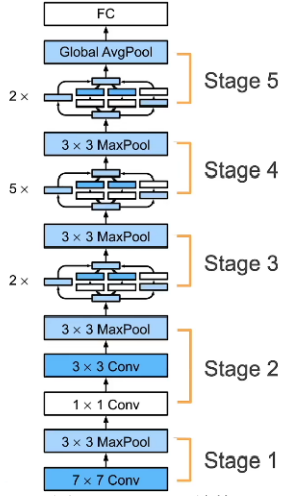


图 3 GoogleNet 结构

后来，Google 公司又调整了 Inception 的结构，加入 batch-normalization[7]、ResNet[8]，替换卷积层，产生了变体 Inception V2、V3[9]、V4[10]。

在我们的分类任务中模型输入层设置为 1 通道，输出层设置为 6 通道，应用了包含 Inception V1 的 GoogLeNet，进行了 20 次迭代，准确率达到 98.4%。

C. ResNet

Kaiming He 在 CVPR2016 提出的 ResNet[8]成为近几年最受欢迎大深度神经网络模型，在多个视觉任务上取得了辉煌战绩。ResNet 解决了深度 CNN 模型难训练的问题，从 14 年的 VGG19 层，到 15 年的 ResNet 多达 152 层，在网络深度不是一个量级，ResNet 除了靠深度取胜，

还有架构上的 trick，这才使得网络的深度发挥出作用，这个 trick 就是残差学习（Residual learning）。下面详细讲述 ResNet 的理论及实现。

从经验来看，网络的深度对模型的性能至关重要，当增加网络层数后，网络可以进行更加复杂的特征模式的提取，所以当模型更深时理论上可以取得更好的结果，但是深度网络也会出现退化问题（Degradation problem）：网络深度增加时，网络准确度出现饱和，甚至出现下降，这使得深度学习模型很难训练。ResNet 提出了残差学习来解决退化问题。对于一个堆积层结构（几层堆积而成）当输入为 x 时其学习到的特征记为 $H(x)$ ，现在我们希望其可以学习到残差 $F(x) = H(x) - x$ ，这样其实原始的学习特征是 $F(x) + x$ 。之所以这样是因为残差学习相比原始特征直接学习更容易。当残差为 0 时，此时堆积层仅仅做了恒等映射，至少网络性能不会下降，实际上残差不会为 0，这也会使得堆积层在输入特征基础上学习到新的特征，从而拥有更好的性能。残差学习的结构如图 4 所示。这有点类似与电路中的“短路”，所以是一种短路连接（shortcut connection）。

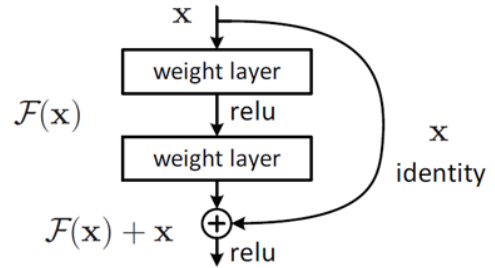


图 4 ResNet 结构

本实验选取了 resnet34 作为分类模型，该模型架构为在 BasicBlock 基础上相继增加了模块数为 3，4，6，3 的 ResBlock，最终在提供的肝脏分类测试集上达到了 99.43% 的准确率。

III. 图像增强

医学影像和 underwater 图像具有很多相似特性，为此我们选取了一种基于自适应参数获取的相对全局直方图拉伸浅海图像增强来增强数据集中的肝脏影像。与自然环境下的图像相比，水下图像更加复杂，图像对比度较差。主要原因有两个方面：

- 水下环境和水浑浊度的影响
- 光的吸收、散射和反射

考虑到水下图像物理模型的有效性和水下图像固有特性的简单性，我们提出了一种新的方法，即相对全局直方图拉伸(RGHS)。它主要是在 RGB 颜色模型中对 G-B 通道均衡化和直方图拉伸的基础上，根据原始图像的分布特征和水下不同波长的光吸收情况来确定拉伸范围参数。双边滤波器是一种简单的、非迭代的边缘保持平滑方法，用于有效地捕捉图像在 RGB 颜色模型中拉伸后的细节。经过对比度校正和颜色校正后，该方法可以保留图像细节，提高图像的可视性。

4. 水下模型

通常使用一个著名的雾霾图像函数模型[11,12]来近似水下散射在背景光下的传播方程。公式如下：

$$I_{\lambda}(x) = J_{\lambda}(x)t_{\lambda}(x) + (1 - t_{\lambda}(x))B_{\lambda}(x) \quad (1)$$

这里光的波长 $\lambda \in \{red, green, blue\}$, x 表示水下图像 $I_\lambda(x)$ 中的像素点, $J_\lambda(x)$ 为 x 点的场景亮度, $t_\lambda(x)$ 为 x 点在水下场景中反射到相机后的剩余能量比, $B_\lambda(x)$ 为背景均匀光。 $J_\lambda(x)t_\lambda(x)$ 描述了场景亮度在水中[13]的直接衰减。注意, 剩余能量比是 λ 和场景-摄像机距离 $d(x)$ 的函数, 它反映了波长在水下距离 $d(x)$ 处的光散射和颜色变化的整体效应。因此, $t_\lambda(x)$ 可以表示为(2), $Nr(\lambda)$ 为归一化剩余能量比, 即每传播单位距离的剩余能量与初始能量之比。绿色和蓝色的光处理更短的波长和高频, 因此衰减比红色对应的非常低。可以将 $Nr(\lambda)$ 对波长的依赖性定义为(3)。

$$t_\lambda(x) = Nr(\lambda)^{d(x)} \quad (2)$$

$$Nr(\lambda) = \begin{cases} 0.8 \sim 0.85 & \text{if } \lambda = 650 \sim 750 \mu\text{m}(\text{red}) \\ 0.93 \sim 0.97 & \text{if } \lambda = 490 \sim 550 \mu\text{m}(\text{green}) \\ 0.95 \sim 0.99 & \text{if } \lambda = 400 \sim 490 \mu\text{m}(\text{blue}) \end{cases} \quad (3)$$

B. 直方图拉伸

由于水下图像分布相对集中, 直方图范围较低, 往往对比度和能见度较低。因此采用直方图拉伸, 使图像通道在整个动态范围内具有更好的像素分布, 从而提高图像对比度。

C. 浅水图像增强模型

在对比度校正步骤中, 对图像应用颜色均衡化和相对全局直方图拉伸(relative global histogram stretch, RGHS)。在保留所需要的彩色水下图像[3]细节的同时, 使用双边滤波器消除上述变换后的噪声。在颜色校正步骤中, 我们应用简单的全局直方图延伸图像的“L”分量, 并在 CIE-Lab 颜色空间(Lab 值分别代表一种颜色的亮度、红绿色度和黄青色度。当 L 值越大, 则表示颜色的亮度越高。 $a > 0$ 时, 表示颜色为红色系, 且 a 值越大表示颜色越红。当 $a < 0$ 时, 则表示颜色属绿色系。 $b > 0$ 为黄色系, $b < 0$ 为系 b 青色系。色泽参数: L 值(Lightness, 亮度), 在 $0 \sim 100$ 之间, 0 表示黑色, 100 表示白色: a 值(Redness, 红色度), 表示红绿之间的色泽, 100 为红色, -80 为绿色; b 值(Yellowness, 黄色度), 表示黄蓝之间的色泽, 100 为黄色, -80 为蓝色。)中调整“ a ”和“ b ”分量。“L”、“ a ”和“ b ”的自适应拉伸可以提高图像的饱和度和亮度, 获得更鲜艳的颜色。

D. 相对全局直方图拉伸 (RGHS)

为了区别于(4)中的全局直方图拉伸, 我们重写了(4)中的相对全局直方图拉伸方程。

$$p_{out} = (p_{in} - I_{in}) \left(\frac{O_{max} - O_{min}}{I_{max} - I_{min}} \right) + O_{min} \quad (4)$$

p_{in} 和 p_{out} 分别是输入和输出像素, $I_{min}, I_{max}, O_{min}, O_{max}$ 分别为拉伸前后图像的自适应参数。下一步, 具体介绍拉伸范围的计算 $[I_{min}, I_{max}]$ 和理想的范围 $[O_{min}, O_{max}]$ 。

拉伸范围的自适应参数获取:

从各种浅水图像的直方图分布中, 我们可以看到 R-G-B 通道的直方图分布类似于定义为(5)的瑞利分布的变化, 这是正值随机变量的连续概率分布。

$$RD = \frac{x}{a^2} e^{-\frac{x^2}{2a^2}} \quad x \geq 0, a > 0 \quad (5)$$

由于水下图像受到各种因素的影响, 为了减少一些极端像素对相对全局直方图拉伸的影响, 如式(6)通常将直方图的拉伸范围取为 $0.1\% \sim 99.9\%$ 。

$$I_{min} = S.sort[S.sort.index(a) * 0.1\%] \quad (6)$$

$$I_{max} = S.sort[-(S.length - S.sort.index(a)) * 0.1\%]$$

其中 S 为每个 R-G-B 通道的图像像素值集合, $S.sort$ 是按升序排列的数据集, $S.sort.index(a)$ 为直方图分布中 a 的索引号, $S.sort[x]$ 表示正序数据集的索引 x 中的值。式(6)的意思是, 从峰分线开始, 将值在总数中最小的 0.1% 的左侧像素和总数中最大的 0.1% 的右侧像素从直方图分布中分离出来, 进行特殊的方法。对于瑞利分布的不同图像和 RGB 通道, I_{min} 和 I_{max} 都是图像和通道敏感的。

为了达到理想的拉伸范围, 我们动态确定每个 R-G-B 通道的最小(O_{min})和最大(O_{max})强度水平值。我们首先用(7)式计算瑞利分布的标准差 σ_λ :

$$\sigma_\lambda = \sqrt{\frac{4-\pi}{2}} a_\lambda = 0.655 a_\lambda, \lambda \in \{R, G, B\} \quad (7)$$

其中 $\lambda \in \{R, G, B\}$ 为 R、G、B 通道, a 为相应通道对应瑞利分布的参数。然后, 我们用式(8)定义了期望范围的最小值 $O_{\lambda min}$:

$$O_{\lambda min} = a_\lambda - \beta_\lambda * \sigma_\lambda, 0 \leq O_{\lambda min} \leq I_{\lambda min} \quad (8)$$

$$\beta_\lambda = \frac{a_\lambda - O_{\lambda min}}{\sigma_\lambda}, \frac{a_\lambda - I_{min}}{\sigma_\lambda} \leq \beta_\lambda \leq \frac{a_\lambda}{\sigma_\lambda} \quad (9)$$

式(9)因为 $a \geq I_{min}$, 可得 $0 \leq \beta_\lambda$ 。由 7 式可得 $\beta_\lambda \leq 1.526$ 。将 $\beta_\lambda = 1$, 式(8)可以被简化为(10)。

$$O_{\lambda min} = a_\lambda - \sigma_\lambda \quad (10)$$

由于水中不同光带的衰减程度不同, 我们必须对 RGB 通道进行单独分析来计算。根据简化的模糊成像模型(1), 通过(11)可以恢复无雾图像 $J_\lambda(x)$ 。

$$J_\lambda(x) = \frac{I_\lambda(x) - (1 - t_\lambda(x))B_\lambda}{kt_\lambda(x)} \quad (11)$$

$k=1.1$ 和 $k=0.9$ 分别为红通道和蓝通道的经验值。最大化恢复的图像 $J_\lambda(x)$ 时, 可以达到所期望的范围 O_{max} 的最大值, 如(12)所示。

$$Max(J_\lambda(x)) = Max \left(\frac{I_\lambda(x) - (1 - t_\lambda(x))B_\lambda}{kt_\lambda(x)} \right) \quad (12)$$

当 $J_\lambda(x)$ 达到最大值, $B_\lambda = 0$, 不同颜色通道的 O_{max} 可由(13)定义为

$$O_{\lambda max} = \frac{I_\lambda}{kt_\lambda} = \frac{a_\lambda + \mu_\lambda * \sigma_\lambda}{k * t_\lambda}, I_{\lambda max} \leq O_{\lambda max} \leq 225 \quad (13)$$

在这里, $t_\lambda(x)$ 用式(2)测量的, R、G、B 通道的 $Nr(\lambda)$ 值分别设为 0.83, 0.95, 0.97 (也见式(3)), $d(x)=3$ 作为场景与相机之间的估计距离。 I_λ 是通道直方图中 a 右侧的一个值, 可以表示为 $a_\lambda + \mu_\lambda * \sigma_\lambda$ 。

由 $O_{\lambda max}$ 的值域值可知, μ_λ 系数满足不等式(14)。

$$\mu_\lambda = \frac{k * t_\lambda * O_{\lambda max} - a_\lambda}{\sigma_\lambda} \quad (14)$$

$$\frac{k * t_\lambda * I_\lambda}{\sigma_\lambda} \leq \mu_\lambda + 1.526 \leq \frac{k * t_\lambda * 255}{\sigma_\lambda}$$

在(14)中, μ_λ 在整数域中没有解或有限解。当 μ_λ 有多个解时, 取其解的平均值; 当 μ_λ 没有解决方案时, 我们简单地设置 $O_{\lambda max} = 255$ 。这些自适应参数既考虑了光的传输特性, 又考虑了原始图像的直方图分布, 避免了图像的过度拉伸和欠拉伸。

E. CIE-Lab 颜色模式下的自适应拉伸

在 RGB 颜色模型中进行对比度校正后, 图像将进行颜色校正。在这个过程中, 水下图像被转换成 CIE-Lab 颜色模型, 以改善颜色表现。在 CIE-Lab 颜色模型中, “L”分量相当于图像的亮度, 表示 $L = 100$ 时最亮的值和 $L = 0$ 时最暗的值。当 $a = 0$ 和 $b = 0$ 时, 颜色通道将呈现真正的中性灰色值。因此, 对 'a' 和 'b' 分量的输出色阶进行修改, 以获得准确的颜色校正, 同时利用 'L' 分量对整个图像的亮度进行调整。首先, 将 CIE-Lab 颜色模型中的浅水图像分解为相应的通道。

“L”分量应用线性滑动拉伸, 由(8)给出, 范围从 0.1%到 99.9%被拉伸到范围[0,100]。图像中上下值的 0.1% 分别设置为 0 和 100。'a'和'b'组件的范围是[-128, 127], 其中 0 是中值。'a'和'b'的延伸而被定义为 s 模型曲线(15)。

$$p_{\chi} = I_{\chi} * \left(\varphi^{1 - \frac{|I_{\chi}|}{128}} \right), \chi \in \{a, b\} \quad (15)$$

式中 I_{χ} 和 p_{χ} 分别为输入和输出像素, $\chi \in \{a, b\}$ 分别为“a”和“b”分量, φ 为最优实验值, 在方法中设为 1.3。式(15)使用指数函数作为拉伸系数, 其值越接近 0, 拉伸越远。

图像的颜色和亮度是图像清晰度和可见性的重要参数。因此, 可以将图像中的物体与背景清晰地区分开来。在 CIE-Lab 颜色模型中对“L”、“a”和“b”三个分量进行自适应拉伸后, 再将通道组合并转换回 RGB 颜色模型。可以生成对比度增强和颜色校正的输出图像作为可感知和可见的最终输出图像。

IV. 实验与结论

本章介绍相关算法的具体实现, 介绍有标注的肝脏超声扫描叶段样本的分类任务以及暗图像增强任务的实现过程。

A. 分类任务的实现

数据集共 11193 张, 人工标记为六种类型, 划分为训练集 (8517 张, 76%) 和测试集 (2676 张, 24%)。利用人工标记的分类训练 AlexNet、GoogLeNet 和 ResNet 三种网络模型, 得到了较好的分类效果。

表 1 数据预处理

1	输入测试集和训练集所在的系统路径
2	获取训练集和测试集下所有图片的系统路径及对应的分类, 形成一个列表
3	使用 SimpleITK 读取所有图片为 PILImage 格式, 后转为 ndarray, 最后转为 tensor
4	使用 torchvision.transforms 进行图像的翻转、剪裁、正则化等变换
5	使用 torch.utils.data.Dataloader 加载每个 batch 的图像和类别列表

表 2 模型的训练和测试

1	按照三个卷积神经网络模型 AlexNet、GoogLeNet、ResNet 的定义, 使用 torch.nn 实现, 参数使用 kaiming 正态分布、常数的方法初始化, 修改输入为单通道、输出为 6 通道
2	将网络模型送入计算设备, 三个网络模型分别进行训练和测试
3	使用 torch.optim.AdamW 优化模型, 进行权重衰减防止模型过拟合, 优化的过程中使用 torch.optim.lr_scheduler 进行学习率的调整
4	使用多分类的交叉熵损失函数
5	每轮迭代遍历全部的训练数据 batch, 计算交叉熵损失的和, 执行梯度下降算法
6	每轮迭代使用训练好的模型参数, 在测试集上计算被模型分类正确的图像数量

B. 增强任务的实现

我们模仿水下图像的处理模型处理超声波图像的光散射等问题。我们采用直方图拉伸方法调整图像像素在 RGB 通道上的分布, 从而提高图像的对比度, 具体的实现步骤如表 3 和表 4 所示。

表 3 相对全局直方图拉伸

1	将图像转化为 png 格式, 读入图片
2	对图像进行归一化处理, 将像素值控制在 0-255 之间
3	将三个通道图像的像素值升序排列, 找出出现次数最多的像素值, 以及出现次数最多的像素值所在最初位置的 0.5%位置, 以及列表右侧对应的位置, 分别作为 a、直方图伸缩的左边界 Imin、直方图伸缩的右边界 Imax
4	出现次数最多的像素值减去标准差, 计算出理想直方图伸缩的左边界 Omin
5	最大化场景像素值, 对出现次数最多的像素值进行一定的增大变换, 计算出右边界 Omax
6	按照公式, 使用 Imax、Imin、Omin、Omax 变换图像三个通道中的每个像素值

表 4 颜色均衡化

1	使用 skimage.color.rgb2lab 方法, 将图像由 RGB 颜色模型表示转为 CIE-Lab 颜色模型表示
2	颜色模型中的 L 分量, 使用直方图拉伸得到的 0.5%—99.5%的像素拉伸到 0—100 的范围
3	使用 s 模型曲线变换的方法, 对颜色模型中的 a、b 分量的像素值进行变换, 拉伸到-128—128 的范围
4	使用 skimage.color.lab2rgb 方法, 将图像由 CIE-Lab 颜色模型模型表示转为 RGB 颜色模型表示
5	使用 cv2.imwrite 将处理后的图像保存为.png 格式

C. 实验设置

分类任务相关的三个卷积神经网络, 都使用 Pytorch 框架在同一 GPU 上训练和测试。模型训练的相关参数设置如下, 迭代次数为 50, batch size 为 8, 初始的学习率为 0.0005, 权重衰减系数为 0.05。训练得到的模型, 按照相同的指标进行模型效果的评价。增强任务使用 Python 语言实现, 代码量不多, 在 CPU 上对随机选取的 25 个图像进行增强, 使用可视化软件查看图像增强效果。

D. 实验结果

实验中使用交叉熵作为分类的损失函数, 使用分类正确率作为模型的评价指标。交叉熵通常作为分类任务的损失函数如式(16), 用来监督模型的训练。其中, $P(i)$ 表示样本的标记属于第*i*个类别的概率, $Q(i)$ 表示模型预测的该样本属于第*i*个类别的概率。

$$H(P, Q) = - \sum_i P(i) \log_2 Q(i) \quad (16)$$

分类模型输出一个样本属于各个类别的概率, 概率值最大的分类作为该样本的模型预测类别。统计标记类别与预测类别相同的样本数, 除以总样本数, 就是模型分类的正确率。分类的正确率介于 0 和 1 之间, 越接近于 1 模型的预测表明越接近于真实的分类。

我们实验中还用到了 mAP(mean Average Precision) 这个指标, 这里首先给出二分类下的混淆矩阵的概念。

	预测的类别 P	预测的类别 N
真实的类别 P	TP	FN
真实的类别 N	FP	TN

精确率 precision 为 $TP/(TP+FP)$ ，召回率 recall 为 $TP/(TP+FN)$ 。精确率和召回率是此消彼长的，常常用所有分类二者形成的曲线下面的面积和除以分类类别数，即 mAP，作为更加合适的评价指标。

我们的实验给出的是一个多分类模型，三个指标的计算方式相比二分类略复杂。对每个样本，模型输出六个通道值经过 softmax 处理为 0-1 之间的概率值，设定一定的阈值来预测样本的类别。以 4 个样本和类别 1 为例，说明精确率、召回率和 mAP 是如何计算的。

样本	1	2	3	4
真实是否为类别 1	0	0	1	1
预测为类别 1 的概率	0.1	0.4	0.35	0.8
阈值 0.35	0	1	1	1
阈值 0.4	0	1	0	1
阈值 0.8	0	0	0	1

在阈值 0.35 下， $TP=2$ ， $FP=1$ ， $TN=1$ ， $FN=0$ ， $precision=0.667$ ， $recall=1$ ；同理，阈值 0.4 下， $precision=0.5$ ， $recall=0.5$ ；同理，阈值 0.8 下， $precision=1$ ， $recall=0.5$ 。最后默认填充一个 $precision=1$ ， $recall=0$ 。得到精确率列表[0.667 0.5 1. 1.]和召回率列表[1. 0.5 0.5 0.]。这样就确定了一条以 precision 为纵坐标、以 recall 为横坐标的曲线。求出每个分类下该曲线下的面积，加和之后除以总分类数，就得到评价指标 mAP。

AlexNet、GoogLeNet、ResNet 三个模型正确率的最优值分别为 98.57%、99.21%、99.43%。说明该分类任务相对简单，结构最简单的卷积神经网络就能达到较好的分类效果，因此更复杂的神经网络对正确率的提升效果不是很明显。

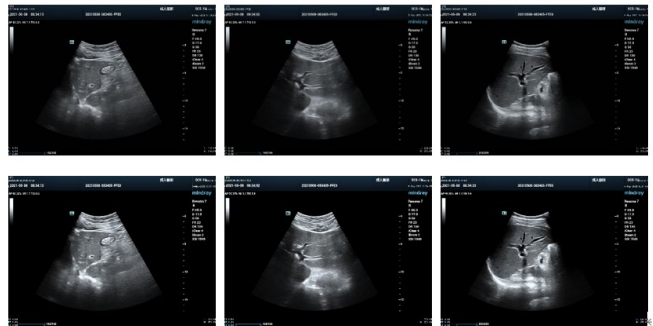


图 5 肝脏图像增强效果，第一行为原图，第二行为对应的增强后图像。

图像增强后，图像的对比度和亮度有明显的改善，颜色更加饱满，甚至能用人眼识别出某些元素。由此得出，将增强水下图像的模型适配到医学超声图像的任务中，也能发挥出一定作用的。

结语

本文使用三种卷积神经网络——AlexNet、GoogLeNet、ResNet，对带标签的图像数据使用交叉熵损失函数进行监督学习，训练和测试之后得到使正确率最佳的模型参数（三个模型的正确率分别为 98.57%、99.21%、99.43%）。为解决在医学样本采样出现的光吸

收现象所造成的图像可视化效果不好的问题，我们将同样解决此问题的水下图像增强的方法（包括相对直方图拉伸和颜色均衡化两个操作）应用到肝脏扫描叶段图像中，提高了图像的对比度，保留了图像的一些细节特征。但是图像的可视化效果没有达到完全人眼可辨别的程度，需要我们的进一步学习来提高图像增强效果。

参考文献

[1] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278-2324.

[2] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

[3] Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1251-1258).

[4] 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*.

[5] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25.

[6] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1-9).

[7] Ioffe, S., & Szegedy, C. (2015, June). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning* (pp. 448-456). PMLR.

[8] 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).

[9] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2818-2826).

[10] Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. A. (2017, February). Inception-v4, inception-resnet and the impact of residual connections on learning. In *Thirty-first AAAI conference on artificial intelligence*.

[11] Schechner, Y. Y., Narasimhan, S. G., & Nayar, S. K. (2001). Instant dehazing of images using polarization. *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*. IEEE.

[12] Narasimhan, S. G., & Nayar, S. K. (2002). Vision and the atmosphere. *International Journal of Computer Vision*, 48(3), 233-254.

[13] Paris, S., & Frédo Durand. (2009). A fast approximation of the bilateral filter using a signal processing approach. *International Journal of Computer Vision*, 81(1), 24-52.

[14] J. Clerk Maxwell, *A Treatise on Electricity and Magnetism*, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68-73.

[15] I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in *Magnetism*, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271-350.

[16] K. Elissa, "Title of paper if known," unpublished.

[17] R. Nicole, "Title of paper with only first word capitalized," *J. Name Stand. Abbrev.*, in press.

[18] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," *IEEE Transl. J. Magn. Japan*, vol. 2, pp. 740-741, August 1987 [Digests 9th Annual Conf. Magnetism Japan, p. 301, 1982].

[19] M. Young, *The Technical Writer's Handbook*. Mill Valley, CA: University Science, 1989.