**A semi-supervised convolutional neural network-based method for steel surface defect recognition**

**摘要**

自动缺陷识别是钢铁生产中的研究热点之一，但是当前的大多数方法都集中在监督学习上，后者依赖于大规模标记的样本。 在某些实际情况下，很难收集和标记足够的样本以进行模型训练，这可能会阻碍大多数最新作品的应用。 使用标记的和未标记的样本进行模型训练的半监督学习可以很好地克服此问题。 本文提出了一种使用卷积神经网络（CNN）的半监督学习方法来识别钢表面缺陷。 所提出的方法需要较少的标记样本，未标记的数据可用于帮助训练。 并且，通过伪标签改善了CNN。 在钢表面缺陷识别基准数据集上的实验结果表明，该方法在有限的标记数据下可以获得良好的性能，达到90.7％的准确度，提高了17.53％。 此外，该方法已应用于某中国钢铁公司的实际案例中，其准确度达到86.72％，明显优于本研讨会中的原始方法。

**1.简介**

质量控制是钢铁生产中的一项艰巨任务。 质量差不仅会造成经济损失，还会极大地影响最终产品的质量[1]。 在缺陷中，表面缺陷是至关重要的问题之一，据现实世界的钢铁车间报告，表面上有超过85％的缺陷发生，这严重影响了最终产品的外观和物理性能。早期，表面缺陷通常是通过人工检查来检查的，但是在工作量大，产量高的情况下，表面缺陷是不稳定且难以实现的[2]。近来，自动表面检查（ASI）系统成为研究热点，并提供了一种快速而强大的方式来代替手动检查[3]。 ASI系统通常包括两个组件：缺陷收集和缺陷识别。第一部分是收集缺陷数据，第二部分是识别缺陷类型[4]。 随着大数据和物联网的发展，缺陷收集在ASI系统中变得很方便，并且大多数研究人员致力于提高识别精度[5]。

在先前的研究中，机器学习（ML）是用于缺陷识别的常规方法之一。 传统的机器学习方法，例如k近邻[5]，支持向量机[6]，学习向量量化[7]和人工神经网络[8]在缺陷识别方面具有良好的性能。 这些方法通常需要显式特征提取，例如直方图[4]，局部二进制模式[9]，小波变换[6]和SIFT [10]。近年来，深度学习（DL）迅速发展。 作为著名的DL模型之一，卷积神经网络（CNN）在识别任务中取得了一些最先进的性能[11-14]。 而且它还广泛应用于与行业相关的任务，例如故障诊断[15,16]和智能电网[17]。 在CNN中，特征是自动学习的，这避免了传统ML方法中显式的特征设计。 受益于此优势，CNN可以在原始图像上工作，而无需进行额外的特征提取，基于CNN的缺陷识别方法越来越受到关注。 Masci等 [18]提出了一种用于表面缺陷的max-pooling卷积神经网络，其准确度为93％。 任等人[19]提出了一种基于Decaf的转移方法，在热轧带钢上达到了99.27％的精度，与以前的工作相比，该精度提高了近10％。 Chen等 [20]提出了一种用于热轧带钢的整体CNN，其准确性为99.89％。

尽管先前的工作在缺陷识别方面有很大的改进，但大多数都集中在有监督的学习上，后者依赖于大规模标记样本进行模型训练。 虽然在某些实际情况下，由于标记大规模样本需要专家知识，所以获取大规模标记样本既困难又昂贵，并且此限制阻止了当前有缺陷识别的应用。此外，由于仅考虑标记的样本，因此大量未标记的样本在ASI系统中处于闲置状态，并且会浪费不必要的存储空间。 使用标签样本和未标签样本进行模型训练的半监督学习提供了解决此问题的另一种方法。 半监督学习需要一些标记的样本进行模型训练，而未标记的样本可用于帮助改善模型性能。 在钢表面缺陷识别中，由于标记数据的成本很高，并且大量未标记的样本处于闲置状态，因此半监督学习更适合此问题。

本文提出了一种使用卷积神经网络（CNN）的半监督学习方法来识别钢表面缺陷。 在这种方法中，通过伪标签（PL）[21]改进了CNN，伪标签是一种有效的半监督框架，可以为未标记的样本生成假标签。所提出的PLCNN的优点包括几个方面。 首先，与监督学习方法相比，提出的PLCNN需要更少的标记样本，从而节省了不必要的样本收集和标记成本。 其次，由于所需的标记样本较少，因此可以更快地部署建议的PLCNN。 最后，闲置的未标记样本可用于模型训练。 基准数据集上的结果表明，所提出的方法在未标记样品的帮助下取得了显着改进，提高了17.53％。 在实际应用中，所提出的方法在有限的标记数据下执行了可接受的结果，与车间中的原始方法相比提高了近50％。 另外，据我们所知，以前从未报道过用于钢表面缺陷识别的半监督学习方法。

本文的其余部分安排如下。 在第2节中详细介绍了所建议的方法，其中简要介绍了CNN，伪标签和所建议的方法。 在第3节中，在基准数据集NEU中对提出的方法进行了测试，以验证该改进。 在第4节中，将所提出的方法应用于真实车间的钢表面缺陷识别案例。 结论和今后的工作在第5节中进行介绍。

**2.用于表面缺陷识别的拟议PLCNN**

本节将介绍拟议PLCNN的详细信息。 在PLCNN中，PseudoLabel（PL）改进了卷积神经网络（CNN），可以在训练过程中使用未标记的数据。

**2.1．卷积神经网络的简介**

作为一种著名的深度学习模型，卷积神经网络（CNN）在识别任务上已经取得了许多最新的表现[13,22,23]。 典型的CNN通常由三部分组成：卷积层，池化层和分类层。 CNN的典型架构如图1所示。在CNN中，卷积层和池化层交替堆叠以进行特征提取。分类层通常连接到最后一层以识别缺陷类型。 每个组件的详细信息将在下面讨论。

**2.1.1．卷积层**

在卷积层中，将卷积运算应用于输入。 假设hl是第L个卷积层的输入，则大小为a \* b的内核通过步长s滑过输入。 令kl是卷积核的数量，表示第i个核的权重和偏差，卷积层的输出可以在等式（1）中定义。其中σ是将输入映射到非线性空间的激活函数，而⊗是卷积算子。



卷积运算后，获得大小为的输出，·为上限函数。

**2.1.2．合并层Pooling layer**

合并层通常紧随卷积层。 在池化层中，应用子采样以减小尺寸并避免过度拟合。 在子采样策略中，由于快速收敛和鲁棒性，通常在CNN中使用最大池化[24]。最大池化的输出是大小为c \* d的不重叠面片的最大值。最大池化产生的输出大小为。

**2.1.3．分类层**

分类层是识别缺陷类型的方法。 通常，分类层位于CNN的末尾。 假设给定分类层C和分类层h的输入，则识别的标签由等式（2）定义：



CNN的目的是使识别的标签和地面真实标签之间的错误最小化。 因此，CNN的损失函数如方程式（3）所示。其中θ是参数集，L（·）表示度量距离，类似于L2-norm。



CNN的训练包含两个步骤：正向传播和反向传播。 在正向传播中，缺陷样本x被馈送到CNN中以计算识别出的标签。在向后传播中，通过最小化识别误差来更新参数集θ。 从（3）中可以明显看出，CNN需要标记数据来优化网络，并且未标记样本不能直接用于CNN。 因此，伪标签被引入到CNN中以解决此问题。

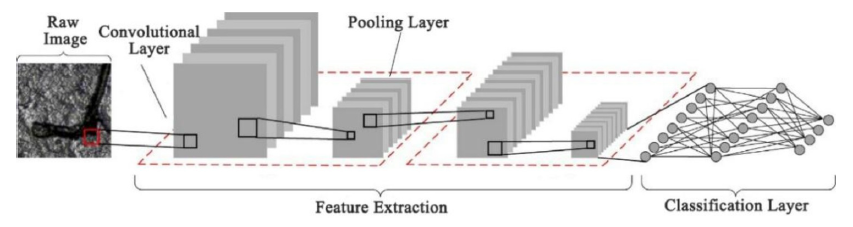


图1. CNN的体系结构。 在CNN中，卷积层和池化层交替堆叠，分类层连接在网络的末端。

**2.2．伪标签**

伪标签（PL）是一种基于自动编码器的简单有效的半监督学习框架[21]。 与有监督的学习不同，PL在训练过程中同时使用标记和未标记的样本。 对于未标记的样本，PL通过识别概率的条件熵生成伪标签，该条件熵度量了识别概率的重叠。 假设每种缺陷类型的识别概率是独立的，则假标签y'由最大识别概率拾取。



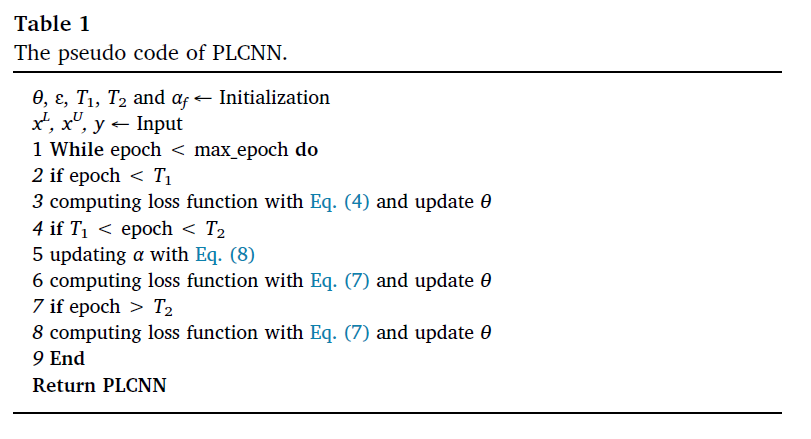
先前的工作表明PL在半监督任务中很有用[25-27]，但是它是由自动编码器提出的。 由于已证明CNN在识别问题上具有更好的性能。 下一节将讨论如何将PL引入CNN。

**2.3． 拟议的PLCNN**

拟议的基于伪标签的CNN（PLCNN）是一个半监督的学习框架，其中标记和未标记的样本用于模型训练。 从这个角度来看，在公式（5）中添加了未标记样本的损失项目。假设带有标记样本xL和未标记样本xU的缺陷集，y〜是PLCNN识别的标记，y和y'分别表示标记样本和未标记样本的假标记的真相，则PLCNN的总体损失函数 可以重写为：



其中，α是权衡系数，用于控制标记样品和未标记样品之间的平衡。 PLCNN的伪代码如表1所示。



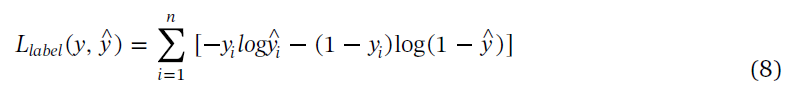
如上所述，PL通常基于自动编码器，它具有对先验知识的无监督预训练。 由于CNN是在没有此先验过程的情况下进行训练的，因此α的调度对于识别结果非常重要。 如果α太大，则会干扰模型，相反，未标记的样本将无法完全利用。 在PLCNN中，对（α）采用（6）中的线性增加调度策略，其中t表示当前迭代。 使用线性增长策略的原因基于两个考虑。 首先，在原始PL中提出了这种线性增加调度策略[21]，并且一些实验结果表明线性增加调度策略取得了良好的性能。 其次，在线性增加中，α平滑地增加，而在其他非线性策略中，例如指数增加，α加速地增加，并且在最后一个时期将大大增加，这可能导致训练过程不稳定。



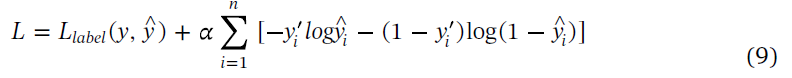
在PLCNN中，培训过程分为两个阶段。 第一阶段是纯监督学习过程，其中α设置为0，并且仅通过标记的样本对模型进行优化。 在第二阶段，α线性增加到常数αf，未标记的数据被输入到模型中。 由于未标记的数据xU需要概率来生成假标签，因此将softmax函数用作分类层，以计算不同标签之间的概率。 在softmax层中，p（y = j | x）的概率由等式（7）计算。其中w是softmax函数中的参数，K是缺陷类型的总数。



为了测量真值标签和PLCNN生成的标签之间的误差，将（8）中的交叉熵定义为损失函数，其中yi和表示第i个样本的原始标签和识别出的标签。



对于，将真实标签y替换为Eq（4）和（7）生成的伪标签y'。总损失函数可以重写为：



在正向传播中，将原始图像前馈以计算识别出的标签。 在反向传播中，参数θ被优化以最小化识别误差。 由于PLCNN的优化很复杂，损失函数是非凸的，因此通常使用（10）中基于梯度的方法来解决这个问题，其中θt表示第t个时期的参数，而ɛ是学习率。 对于不同层之间的传播，导数由链规则计算。



在文献[18]中，已经提出了一种基于CNN的缺陷识别方法，但是这两种方法之间存在一些差异。 首先，提出的PLCNN是一种半监督方法，将标记和未标记的样本都用于模型训练。文献[18]中的方法是一种监督方法，只能使用标记的样品。 其次，这两种方法具有不同的网络体系结构。

2.4．作为数据增强的dropout

缺陷识别中训练集的规模小于其他识别任务[13]。 而且过度拟合是不可避免的，因为该模型在训练集上表现良好，而在测试时表现不佳。 这个问题在PLCNN的第一阶段很明显，即仅使用标记的样本。 为防止此问题，在分类层中使用dropout [28]作为数据增强。 在辍学层中，神经单元以逐层概率随机关闭。辍学的定义在（11）中显示。 随机关闭类似于添加一些噪声以改善变化的方式。 先前的工作已经证明，辍学可以被视为数据的扩充，可以有效地防止过度拟合[28]。

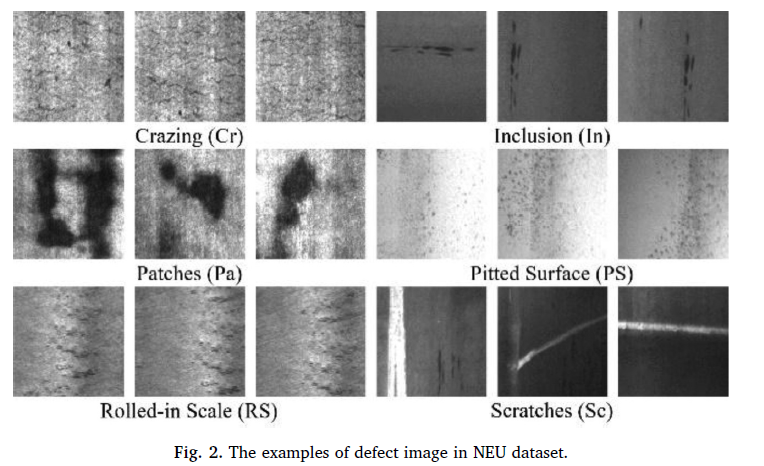


**3.在基准数据集上PLCNN的实验结果**

在本节中，将在基准数据集NEU上测试所提出的方法以评估改进。 介绍了实验细节和参数设置。 结果表明，所提出的方法具有改进的性能，并且未标记的数据是有用的。

**3.1．数据描述**

东北大学数据集（NEU）[9]是热轧钢带基准数据集。该基准数据集包含六个典型的钢表面缺陷，包括开裂（Cr），夹杂物（In），斑块（Pa），点蚀表面（PS），轧入氧化皮（RS）和划痕（Sc）。 对于每个缺陷，它包含300张带标记的灰度图像，原始分辨率为200 \*200。NEU的示例如图2所示。同一列中的缺陷属于类内缺陷，而不同列则属于内部缺陷。 类间缺陷。 在该数据集中，将相同的缺陷定义为类间缺陷，而将不同类型的缺陷定义为类间缺陷。 如图2所示，显然类别间的差异很小，但是类别内的差异却是多种多样的。 在此实验中，NEU数据集分为训练集和测试集。 训练集包含1500个样本，每个缺陷会保留50个带有真值标签的样本，并删除剩余的标签。



为NEU数据集建立的拟议PLCNN模型具有6层，其架构如图3所示。在PLCNN中，第一卷积层具有16个5×5卷积核，其余每层具有32个2×2核。 最大池用于加速池层的收敛。 在CNN的末端，在softmax层之前使用完全连接层和辍学层。 另外，将辍学率pdr设定为0.5，将Eq.10中学习率ɛ设定为设置为0.001。 通过交叉验证选择所有超参数，表2给出了PLCNN的摘要。为了确保公平，所有方法都对原始数据起作用，而无需进行额外的特征提取。 这些方法在python工作站上运行10次。 并在TensorFlow上通过GPU加速了该方法。

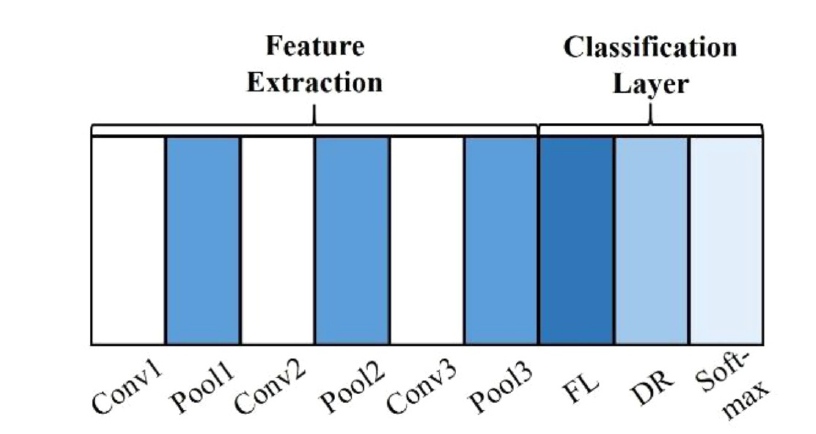
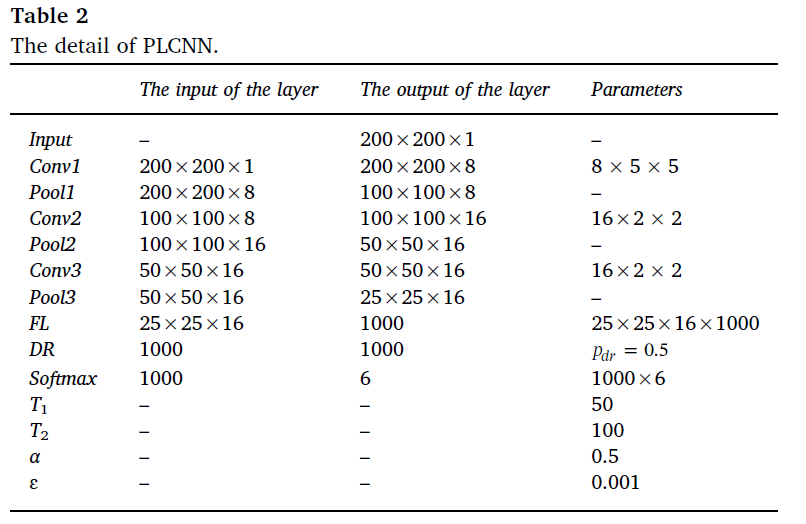


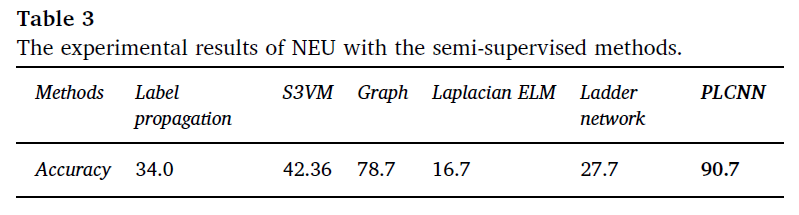
图3. NEU数据集的PLCNN架构。 Conv：卷积层，Pool：池化层，FL：全连接层，DR：辍学层。



应当注意，所有这些方法都是在原始图像上实现的。 这是因为深度学习可以自动学习功能，从而避免了明确的功能设计。 此外，额外的特征提取将需要更多的计算，并且会影响响应时间。

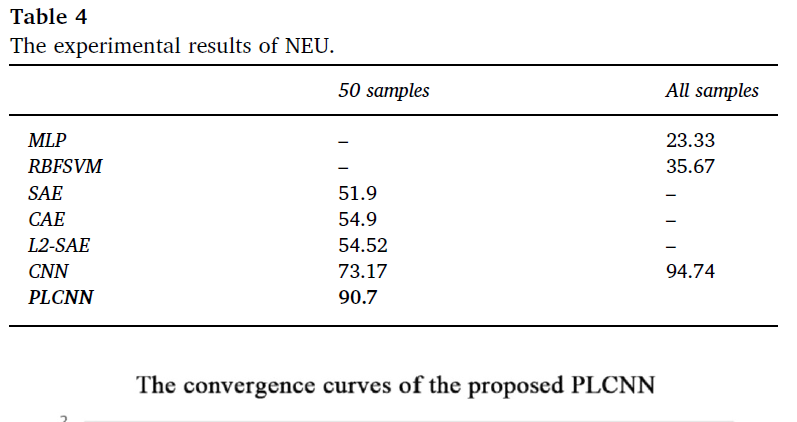
**3.2．实验结果**

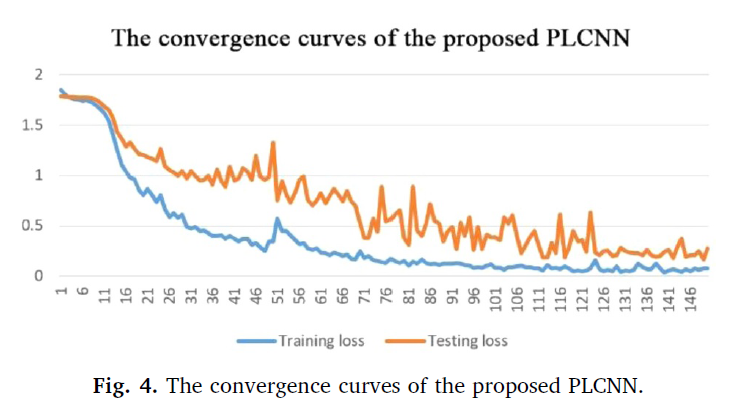
在本节中，实验包含两个部分。 在第一部分中，将提出的PLCNN与常规的半监督方法进行比较，第二部分是与常规的缺陷识别方法进行比较。 第一部分中的比较方法包括标签传播[29]，S3VM [30]，图[31]，拉普拉斯ELM [32]和阶梯网络[33]。 本部分旨在评估拟议的PLCNN在半监督学习中的性能，其结果如表3所示。



从该结果可以看出，所提出的PLCNN的性能优于其他半监督方法，其准确性比比较方法中的最佳结果提高了12％。 比较方法的准确性为34.0％，42.36％，78.7％，16.7％和27.7％。

在第二部分中，将拟议的PLCNN与常规的深度学习和缺陷识别方法（包括多层感知器（MLP），径向基函数支持向量机（RBF-SVM），堆叠式自动编码器[34]（SAE），具有L2调节的变型自动编码器（L2-SAE），压缩式自动编码器[35]（CAE）和常规的CNN。 实验结果示于表4，收敛曲线示于图4。

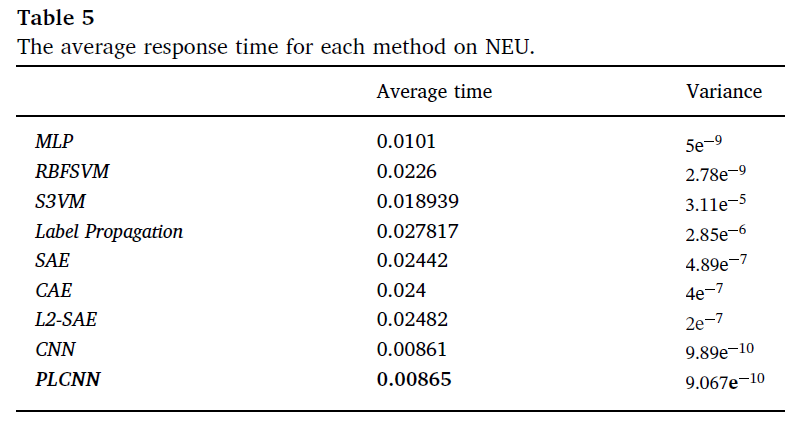




结果表明，所提出的PLCNN具有改进的性能。与其他方法相比，该方法的准确率达到了90.7％。 SAE，CAE，L2-SAE和CNN的其余识别准确度分别为51.9％，54.9％，54.52％和73.17％。这表明PLCNN的性能优于其他。 不去除标签，MLP和RBF-SVM的准确度分别为23.33％和35.67％。结果表明，传统方法在该问题上是无效的。没有特征提取，这些方法将无法获得预期的结果。 此外，CNN和提议的PLCNN的结果表明，未标记的样本对于提高识别性能可能是有用的。

收敛曲线如图4所示。从曲线来看，训练损失正在平稳下降，并且在时期50中这两条曲线都增加了。这是因为未标记的样本加入了训练过程。

除精度外，响应时间是钢铁生产中的另一个重要指标。 为了评估该方法的响应时间，将原始图像馈入到最小批处理大小为10的不同模型中。每张图像的平均响应时间如表5所示。得益于GPU的加速，该响应 所提出的方法的时间已被证明是迅速而可靠的。



**3.3．讨论**

如表3和表4所示，该方法取得了良好的性能。 借助未标记的样品，该方法相对于常规CNN改进了17.53％。 与传统的机器学习方法相比，所提出的PLCNN具有重要的意义。 首先，在PLCNN中，使用标记样本的监督学习阶段是为下一阶段学习先验知识。 一旦合并了未标记的数据，此模型将尝试查找与假标签最相似的缺陷类型。 其次，假标签也可以被视为防止模型仅从标签样本中学习的规则。 由于在阶段2的早期阶段伪造标签的置信度低，因此系数α从较小的值开始增加。 此外，应该注意的是，αf最终固定为0.5，而不是1，这表明假标签不能视为真标签。

**4. PLCNN在表面检验真实案例中的应用**

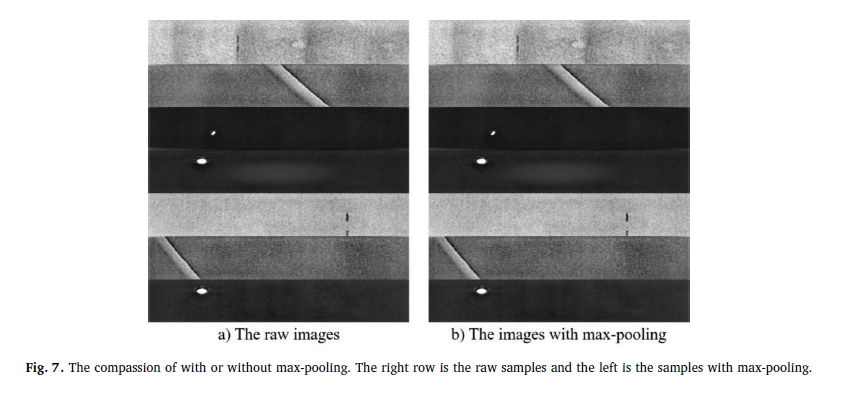
在本节中，将所提出的方法用于解决中国一家钢厂的真实表面缺陷识别案例。 在这个车间中，建立了一个自动的表面检查系统，收集了一些缺陷图像并应用了一些识别方法，但是它们都没有按预期工作。 这种情况的详细情况和要求如下所述。

**4.1．案例描述**

该案例来自一个现实世界的冷轧钢车间，该车间生产用于车辆的电镀锌钢带。 生产流程图如图5所示。根据该车间的统计，该产品的合格率仅为58.6％，并且表面上出现95.4％以上的缺陷。 手动识别是准确的，但不能满足高速生产。 在该车间中，建立了自动表面检查（ASI）系统，并收集了分辨率为1600×240的缺陷图像。 在早期检查中，使用支持向量机（SVM）来识别缺陷，但是识别精度低于30％，在此车间中至少需要85％。 为了提高识别精度，提供了具有6个典型缺陷的2500个缺陷图像，并手动标记了其中300个图像，即滚动痕迹，条子，边缘缺陷，划痕和孔眼。 此外，由于图像较宽，一幅图像可能包含多个缺陷。 图像中出现条子和孔洞的典型缺陷被添加到缺陷类型中。 收集的图像示例如图6所示。

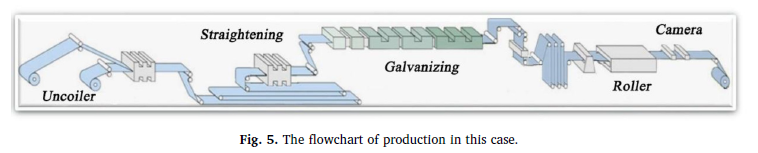
**4.2．用于实际钢表面缺陷识别的PLCNN**

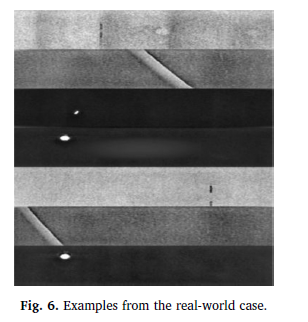
受可计算性限制，缺陷样本的尺寸太大，无法直接输入PLCNN。 传统的降维方法（如PCA和ICA）通常假定测量值的多元正态性，这限制了在实际问题中的应用[36]。 因此，在第一卷积层之前采用2×2最大池化层以减小尺寸。 带有或不带有最大合并的图像显示在图7中。缺陷特征保留，并且仅具有原始图像的四分之一尺寸。



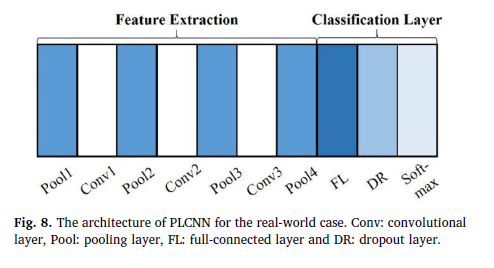


与上述实验相同，收集的数据分为训练集和测试集。 训练集有300个标记的样本和1000个未标记的样本是从收集的数据中随机选择的。 测试集包含从其余图像中选择的500个样本。 在模型训练之前，应用（12）中的归一化以避免幅度大的点主导随后的计算。

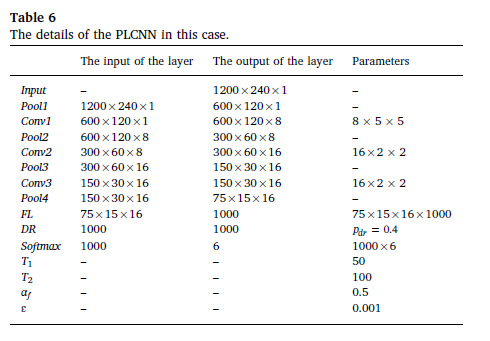




在这种情况下，PLCNN具有7个隐藏层：1）2×2的最大合并层。 2）具有8个5×5滤镜的卷积层。 3）2×2的最大合并层。 4）具有16个2×2滤波器的卷积层。 5）2×2的最大池化层。 6）具有16个2×2滤波器的卷积层。 7）2×2的最大池化层。 最后，将softmax连接为分类层。 由于训练集的大小较小，因此在分类层中添加了作为数据增强的退出层。 pdr设置为具有二进制分布的0.4。 在这种情况下，此方法的体系结构如图8所示。

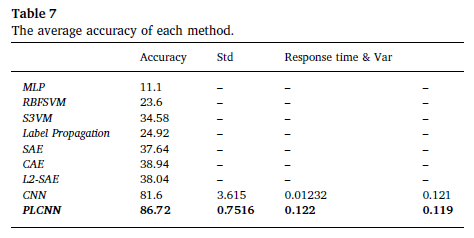


ReLU在（1）中用作激活函数，具有很强的非线性和稀疏表示能力。 通过随机梯度下降优化学习过程，学习率= 0.001，批次大小为10。T1，T2和αf分别为50、100和0.5。 由于此实验需要人工检查以验证识别结果，因此使用交叉验证或其他策略选择超参数的成本很高。 在这种情况下，PLCNN的设置遵循第3节中的结果。表6汇总了所有设置和超参数。该实验运行了5次，每次训练150个纪元。



**4.3．应用结果**

在本案例研究中，将使用第3节中提到的所有方法进行比较。 表7列出了识别精度。



根据比较结果，PLCNN的准确度为86.72％，与原始方法相比提高了近50％，在本研讨会中是可以接受的。 MLP，RBF-SVM，S3VM，标签传播，SAE，CAE，L2-SAE和CNN的识别准确度分别为11.1％，23.6％，34.58％，24.92％，37.64％，38.96％，38.04％和81.6％。 该结果表明，如果没有额外的特征提取，大多数比较方法将无法正常工作。 由于在这种情况下只有PLCNN和CNN有效，因此表7还列出了这两种方法的响应时间。 为了评估响应时间，给模型填充一个小批量，并通过每个图像的平均时间来计算响应时间。 结果表明，PLCNN随GPU的加速而迅速，并且响应时间和var也可以接受。

图9和图10是PLCNN和CNN的混淆矩阵。 从图9中，提出的方法可以区分辊痕，边缘缺陷以及辊痕和条子同时出现的特殊情况。 在本实验中，只有提出的方法可以区分这种特殊情况。 对于其他缺陷，所提出的方法也进行了重大改进。 与CNN相比，该方法在条子和刮擦方面具有改进的性能。 但是这两种方法都无法识别漏洞。 对于此问题，孔和划痕之间的相似性很小，因此很容易错误分类。这与缺陷条和边缘缺陷的原因相同。

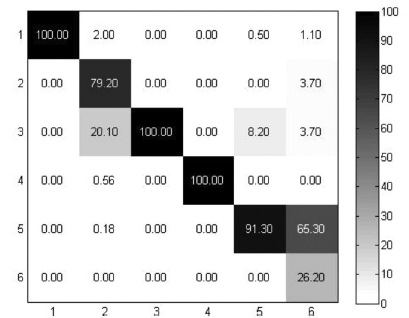


图9. PLCNN的混淆矩阵 数字表示缺陷类型。 1：卷痕，2：细条，3：边缘缺陷，4：细条+卷痕，5：划痕6：孔。 x轴是真相标签，Y轴是公认的标签。

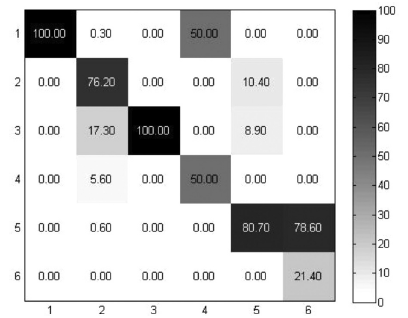


图10. CNN的混淆矩阵。 数字表示缺陷类型。 1：卷痕，2：细条，3：边缘缺陷，4：细条+卷痕，5：划痕6：孔。 x轴是真相标签，Y轴是公认的标签。

**4.4．实施细节和讨论**

尽管所提出的方法满足了本次研讨会的要求，但其识别率显然低于监督学习方法，并且比第3节中的基准数据集的结果差。有很多原因。 首先，计算能力限制了结果，随着网络变得越来越深，越来越复杂，识别率将提高。 其次，无论是否标记，训练集仍然很小，无法优化PLCNN。 收集的样本越多，提供的结果越好。 最后，从实际案例中收​​集的图像质量不稳定，因此很难识别。 拟议中的PLCNN在具有八核，16 GB内存和GTX1060 Nvidia GPU的台式机上实现。 该实验基于Python，并且软件包涉及Keras [37]，Scikit-Learn [38]和NumPy [39]。

**5.结论和未来研究**

本文提出了一种使用CNN的半监督学习方法来识别钢表面缺陷。 所提出的方法需要较少的标记样本，未标记的数据可用于帮助训练。 实验结果表明，该方法在未标记样品的帮助下具有明显的改进。 与传统的依赖于大规模标记样本的方法相比，该方法使用了空闲的未标记数据，节省了数据标记的成本，可以更快地部署，并且更适合于标记样本有限的缺陷识别任务。 在PLCNN的应用中，尽管未标记的数据在训练时很有用，但应注意标记的样本仍然很重要，如果在生产过程中收集到更多的标记数据，则应将其添加到此模型中以改善识别结果。

PLCNN的局限性有以下几个方面。 首先，与监督学习方法相比，PLCNN的精度较低，需要在生产中不断提高。其次，在早期生产中，训练集的规模是有限的，如果在训练过程中未学习到新出现的缺陷类型，则会被错误分类。 因此，未来的工作可以集中在两个方面。 一种将通过半监督学习方法来提高识别率。 另一个是将所提出的方法与不断增加的学习相结合，这更适合于新出现的缺陷类型。