计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2020,29(7):180−185 [doi: 10.15888/j.cnki.csa.007487]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于深度学习高分辨率遥感影像语义分割①

尚群锋, 沈　炜, 帅世渊

(浙江理工大学 信息学院, 杭州 310018)
通讯作者: 尚群锋, E-mail: nianxiongdi1231@163.com

摘　要: 高分辨率遥感影像含有丰富的地理信息. 目前基于传统神经网络的语义分割模型不能够对遥感影像中小物体进行更高维度的特征提取, 导致分割错误率较高. 本文提出一种基于编码与解码结构特征连接的方法, 对DeconvNet 网络模型进行改进. 模型在编码时, 通过记录池化索引的位置并应用于上池化中, 能够保留空间结构信息; 在解码时, 利用编码与解码对应特征层连接的方式使模型有效地进行特征提取. 在模型训练时, 使用设计的预训练模型, 可以有效地扩充数据, 来解决模型的过拟合问题. 实验结果表明, 在对优化器、学习率和损失函数适当调整的基础上, 使用扩充后的数据集进行训练, 对遥感影像验证集的分割精确度达到 95% 左右, 相对于 DeconvNet 和 UNet 网络模型分割精确度有显著提升.

关键词: 深度学习; 语义分割; 遥感影像; 反卷积网络

引用格式: 尚群锋,沈炜,帅世渊.基于深度学习高分辨率遥感影像语义分割.计算机系统应用,2020,29(7):180–185. http://www.c-s-a.org.cn/1003-3254/7487.html

Semantic Segmentation of High Resolution Remote Sensing Image Based on Deep Learning

SHANG Qun-Feng, SHEN Wei, SHUAI Shi-Yuan

(School of Information Science and Technology, Zhejiang Sci-Tech University, Hangzhou 310018, China)

Abstract: High-resolution remote sensing images contains rich geographic information. At present, the semantic segmentation model based on the traditional neural network cannot extract the features of small and medium-sized objects in remote sensing images, resulting in high segmentation error rate. This study proposes a method based on the connection of encoder and decoder structure features to improve the DeconvNet network model. The model can retain the spatial structure information by recording the location of the pool index and applying it to the upper pool when being encoded. During decoding, the model can effectively extract features by connecting the corresponding feature layer of encoder and decoder. During model training, the pre-training model designed can effectively expand the data to solve the problem of model over-fitting. The experimental results show that, based on the proper adjustment of optimizer, learning rate and loss function, the accuracy of remote sensing images semantic segmentation in the validation database is about 95% by using the extended dataset for training, which is significantly improved compared with the DeconvNet and UNet network models.

Key words: deep learning; semantic segmentation; remote sensing image; deconvolution network

遥感技术 (remote sensing technique) 是指通过自身辐射或反射的电磁波、可见光等对物体进行探测与识别. 根据该技术所提供信息生成的地表遥感影像图, 被广泛应用于自然灾害检测[1]、城市规划[2]与土地覆盖检测[3]等领域. 通过对地表纹理、位置、阴影、形状、大小与物体空间位置等信息细致地观测, 高分辨率遥

① 收稿时间: 2019-12-01; 修改时间: 2020-01-03; 采用时间: 2020-01-07; csa 在线出版时间: 2020-07-03

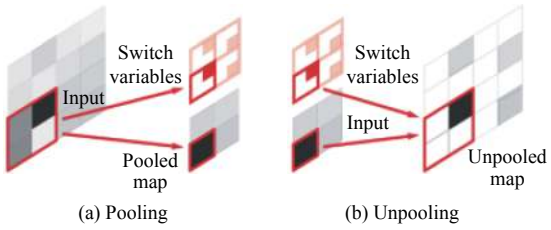
感影像可以清楚地表达表纹理与空间信息的特征, 为提取城市信息提供服务, 因此探索深度学习语义分割网络模型在高分辨率遥感影像中的应用具有重要研究价值与意义.

高分辨率遥感影像含有丰富的地理信息, 利用传统机器学习对遥感影像特征分析与特征提取的过程都具有一定复杂性, 不能够有效地对空间结构与物体边缘特征进行提取, 有显著的局限性. 而深度学习网络模型有学习能力强、覆盖范围广与可移植性好等优点, 能够通过神经网络模型对空间特征与物体边缘特征进行提取, 可以有效解决上述问题, 使得深度学习对遥感影像语义分割的研究成为焦点, 有助于改善高分辨率遥感影像的分割精度. 潘等人[4]采用随机森林回归机器学习算法对建筑物进行提取, 从视觉显著性的角度分进行分析, 可以有效地提取建筑物; 黄等人[5]提出改进迭代条件模型的遥感影像语义分割方法, 首先对遥感影像使用 L0 梯度最小化模型去噪, 然后使用迭代条件模型, 再更新遥感影像中每个点的标记, 实现对遥感影像语义分割, 提升了物体整体的分割效果, 但是得到的遥感影像还是存在椒盐问题; 李等人[6]提出基于深度残差网络的高分辨率遥感影像分割方法, 该方法通过运用深度残差与全卷积网络进行端到端语义分割模型的构建, 并在全卷积网络中引入空洞卷积, 解决分割粗糙问题, 在小尺度对象上也具有较好的分割效果. 苏等人[7]提出一种基于 UNet 改进的深度卷积神经网络, 分为预训练阶段与训练阶段. 预训练阶段对以影像进行翻转、色彩调整与加噪等处理, 实现对数据集的扩充. 训练阶段把每个类地物转换为二分类模型进行目标训练, 随后将各个分类子图进行连接生成最终的语义分割图像, 采用集成学习的策略提高分割精度. 这些方法在一定程度上都提升了分割精度, 但存在着一些不足之处, 例如空间信息丢失, 细节信息被忽略等问题, 而反卷积神经网络模型[8]可以利用上池化与反卷积结构有效解决上述问题, 优化模型的分割效果. 然而反卷积网络对预测小物体分割仍存在分割精度不足, 需进一步研究与探索.

针对上述问题, 本文提出新的方法, 利用编码结构与解码结构对应特征层连接的方式与数据预训练模型对反卷积网络模型优化, 有效对遥感影像物体空间信息、边缘信息与像素之间的关系进行处理, 提高了模型的分割精度.

1 反卷积网络

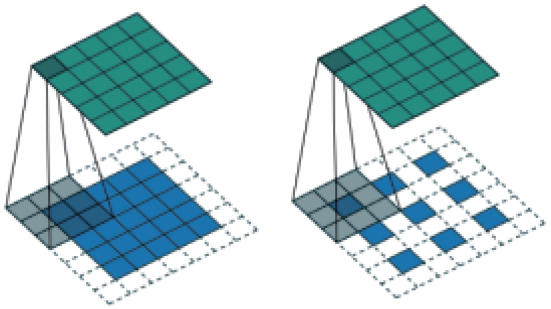
反卷积网络 (Deconvolutional Networks, DeconvNet) 语义分割模型分为编码与解码结构, 且对称. 在编码与解码过程中, 分别进行下采样与上采样, 如图 1 为下采样与上采样结构, 池化 (Pooling) 代表下采样, 上池化 (Unpooling) 代表上采样. 池化可以通过提取单一的值对当前感受野去噪, 且保留有效的信息, 有助于分类. 但是在池化过程中, 其空间信息会丢失, 导致预测不精确. 对于这个问题, 模型中运用上池化层和反卷积层来防止空间信息的丢失. 池化层使用 switch variables 记录池化索引的位置, 在上池化中利用 switch variables 索引记录的位置, 映射到对应的空间位置来保留空间信息, 从而提升图像分割精度.



(a) Pooling (b) Unpooling

图 1 下采样与上采样结构

在上池化产生稀疏特征图, 反卷积通过多个滤波器学习[9], 使最终结果更加的准确. 图 2 为反卷积两种实现方式[10], 图 2(a) 表示边缘填充零, 图 2(b) 表示间隔填充零.



(a) 边缘填充零 (b) 间隔填充零

图 2 反卷积的方式

2 改进反卷积网络

本文根据文献[11]中特征层连接的方法, 提出了改进反卷积网络模型, 并将模型分为编码与解码两部分: 编码部分由 5 组结构组成, 且均含有卷积层、BN 层与池化层; 解码部分也由 5 组结构组成, 都含有上池化

层、反卷积层、BN 层与设计的连接层. 改进反卷积网络模型的结构如图 3 所示, 图中矩阵框与下方数字表示特征图和其维度. 黑色箭头表示卷积+BN+ReLU、池化、上池化、反卷积+BN 以及反卷积+Softmax 等操作.下文将详细介绍网络的各层结构.

2.1 预处理

由于原始遥感图像偏大, 直接把原始图像输入到神经网络中, 需要消耗大量的内存和显存, 所以要进行预处理后作为输入层, 本文选择的预处理后遥感图像尺寸为 512×512 像素三通道图像.

2.2 编码结构

对于编码结构的设计, 本文采用 5 组结构大致相同, 参数略有不同的卷积, 每组包含多个卷积层. 以第 1 组编码结构 (第 2 层到第 3 层) 为例进行详细介绍. 第 2 层使用 64 组 3×3×3 的卷积核, 在卷积时, 对输入层图像进行上下左右边界进行像素扩充, 扩充像素值设置为 0, 卷积的步长设置为 1, 卷积后产生 64 个特征映射图. 然而卷积操作是线性操作, 为了让模型非线形拟合的能力, 本文使用的激活函数为 ReLU (Rectified Linear Units), 来增加模型的非线性能力:

$$f(x) = \max(0, x) \tag{1}$$

在每次卷积后, 还需要对得到的特征图进行 BN[12](Batch Normalization) 操作, 目的是对特征图进行归一化处理. 第 3 层的卷积和前一层类似, 不同的卷积层可以提取不同特征. 表 1 中列举第 1 组编码结构参数的设置.

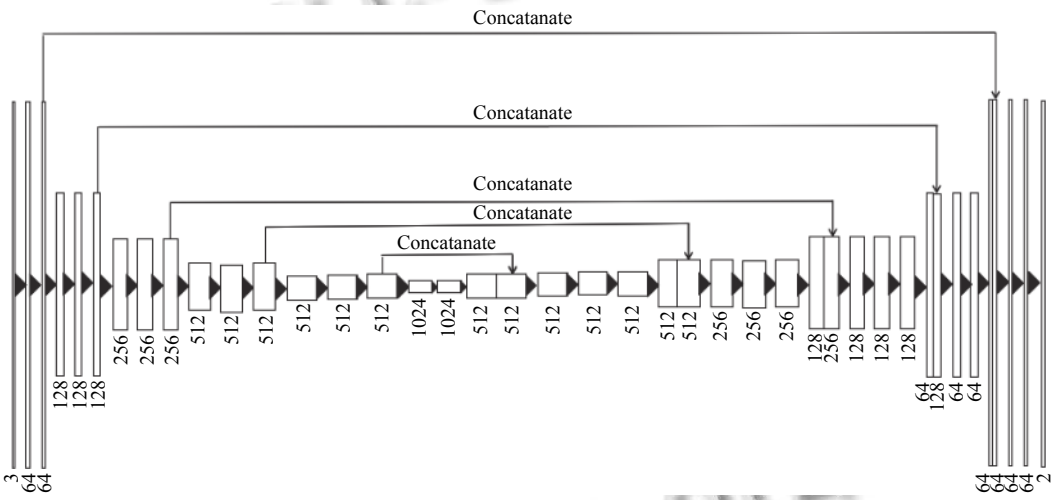


图 3 改进反卷积网络模型结构

表 1 编码第一组结构

层数	输入	卷积核	输出
第 2 层	512×512×3	64 组 3×3×3	512×512×64
第 3 层	512×512×64	128 组 3×3×64	512×512×128

第 4 层是池化层. 由于卷积产生的特征图将会占用大量内存空间, 导致模型训练困难, 可以利用池化操作来解决这个问题, 不仅降低特征向量的维度, 而且有效防止过拟合. 池化层采用 2×2 的核, 步长为 2 最大池化, 即每 4 个相邻的像素中保留最大的值, 将卷积层的特征图尺寸缩小为 0.25 倍.

其他 4 组的编码结构和第一组类似, 依次进行卷积与池化操作, 第 5 组最后得到 512 个 64×64 个特征图. 在第 5 组末尾面增加两个卷积层, 进一步进行特征的提取, 第一个卷积层使用的是 7×7 的卷积核, 产生 1024 个特征图. 第 2 个卷积层使用 1×1 的卷积核, 最终产生 1024 个 10×10 特征图.

2.3 解码结构

解码结构目的是把特征图还原为原始尺寸, 通过上池化、反卷积与设计的连接层, 能够对空间信息与物体边缘特图有效提取. 在反卷积的过程中, 设置卷积的滑动窗口大小为 3×3, 步长为 1. 解码结构首先是一个反卷积层与 5 组反卷积结构组成, 每组对应多个反卷积层、一个上池化层与一个连接层. 这一个反卷积层产生 512 个 16×16 特征图. 其余 5 组反卷积结构, 以第一组解码结构为例进行详细地介绍. 首先对 512 个 16×16 特征图进行上池化, 得到 512 个 32×32 特征图,

然后与编码结构第五组最后一层卷积进行特征连接, 此时特征图变为 1024 个, 经过 3 次反卷积操作之后, 得到 512 个 32×32 密集的特征图. 接着对第二组解码结构进行相同的操作, 以此类推. 表 2 为解码结构每组最后一个反卷积层的尺寸.

表 2 解码结构各组尺寸

组数	尺寸
第 1 组	32×32×512
第 2 组	64×64×256
第 3 组	128×128×128
第 4 组	256×256×64
第 5 组	512×512×64

最后通过反卷积操作, 得到 512×512×2 的特征图像.

2.4 网络模型的参数设置

对于建筑物像素级别分类, 属于二分类任务. 通过 Softmax 将每一个像素点值映射为一个概率值, 从而确定每一个像素值是否属于建筑物类, 达到像素级别的预测, 即:

$$p_i = \frac{e^{y_i}}{\sum_{j=1}^{2} e^{y_i}} \tag{2}$$

其中, p_i代表当前分类的概率值, y_i代表当前预测的值.

损失函数选择是非常关键的, 它是 BP 算法[13]的核心部分, 也是用来衡量预测值和标签值之间的差异, 通过计算误差, 将得到的误差进行反向传播. 网络模型在训练过程中使用交叉熵[14]作为损失函数, 可以有效地对模型进行预测, 交叉熵的损失函数如下:

$$H = -\frac{1}{n}\sum_{i=1}^{n}(y_i \log \widehat{y}_i + (1-y_i)\log(1-\widehat{y}_i)) \tag{3}$$

其中, n代表类别, y_i代表 label 的真实值, 代表为$\widehat{y}_i$预测值.

3 结果与分析

3.1 实验环境

本次实验采取 ubuntu 下的 keras 深度学习框架, 使用 Pycharm 软件进行程序编写, 实验平台的配置如下: CPU 为 Intel(R) Pentium(R) CPU G3260 @ 3.30 GHz, GPU 为 NVIDIA GTX 6G, 1 TB 硬盘.

3.2 数据预训练模型

数据集是马萨诸塞州建筑物拍摄的遥感影像[15], 每幅图像的大小为 1500×1500 像素, 数据集覆盖了城市、郊区和农村地区, 面积超过 2600 平方公里. 由于原始遥感影像较大, 需要对图像进行预处理, 产生符合网络输入的大小. 这里运用了对数据集进行动态扩充的方法, 动态地进行训练. 在训练时, 通过随机 scale 的方式, 进行动态加载, 让模型消除欠拟合或过拟合. 图 4 为数据预处理过程, 首先对原始遥感影进行随机的 scale, 再进行随机翻转或旋转遥感影像, 这时得到的遥感影像不符合网络模型的输入图像大小, 需要对图像进行裁剪, 得到符合模型的输入图像. 然而在每次迭代结束时, 就需要重新获取输入图像, 从原始遥感影像到输入图像的过程, 实现对数据集的扩充. 这样做的目的是为了不断有新的数据加载到模型中, 提高模型的分割效果.

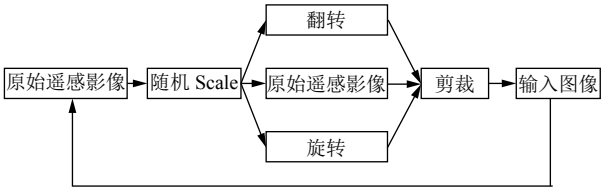


图 4 数据的预处理

3.3 评价指标

评价指标是用来评估不同算法在某一方面的效果是否最佳, 这样可以帮我们对算法进行不同程度的优化. 为定量分析图像语义分割精度, 本文采用像素精确度、召回率、准确率来对模型评估.

像素精确度定义为预测正负样本正确个数的和总样本个数的比值.

$$ACC = \frac{TP+TN}{TP+TN+FN+FP} \tag{4}$$

召回率定义为预测正样本正确个数和标注图像中正样本总数的比值.

$$R = \frac{TP}{TP+FN} \tag{5}$$

准确度定义为预测正样本正确的个数和预测为正样本的总数的比值.

$$P = \frac{TP}{TP+FP} \tag{6}$$

$F1$定义为召回率和准确率的调和均值, 是综合考虑召回率和准确率的指标.

$$F1 = \frac{2\times(R\times P)}{R+P} \tag{7}$$

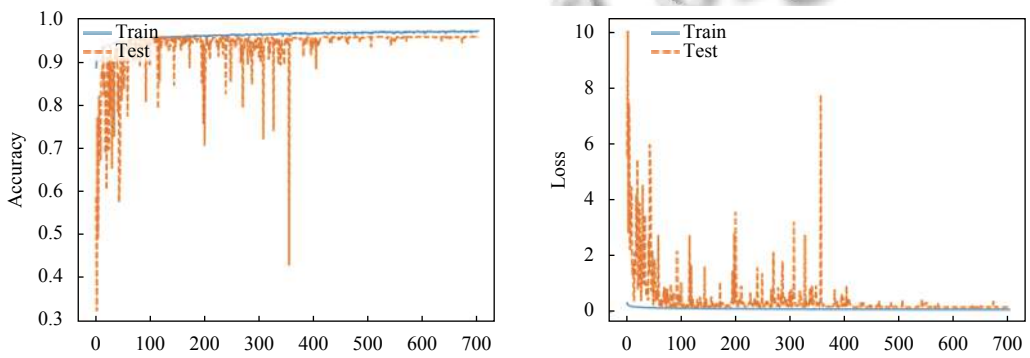
上述公式中, $F1$为真正例, 代表的是正样本预测结果为正确的数目; FN为假反例, 代表的是负样本被预测为正样本的数目; TP为假正例, 代表的是正样本被预测为负样本的数目; TN为真反例, 代表的是负样本预测结果为正确的数目.

3.4 结果分析

如图 5 曲线图代表使用改进的反卷积网络模型在马萨诸塞州建筑物拍摄的遥感影像数据集进行训练的像素精确度和损失函数曲线, 图 5(a) 代表的是像素精确度, 在模型训练到 400 次 epoch 的时候, 训练集上的精确度为 96% 左右, 测试集上的精确度达到 95% 左右, 达到良好的效果; 图 5(b) 代表的是损失函数曲线, 模型在训练 400 次左右的时候趋向于平稳且在 0.1 附近.

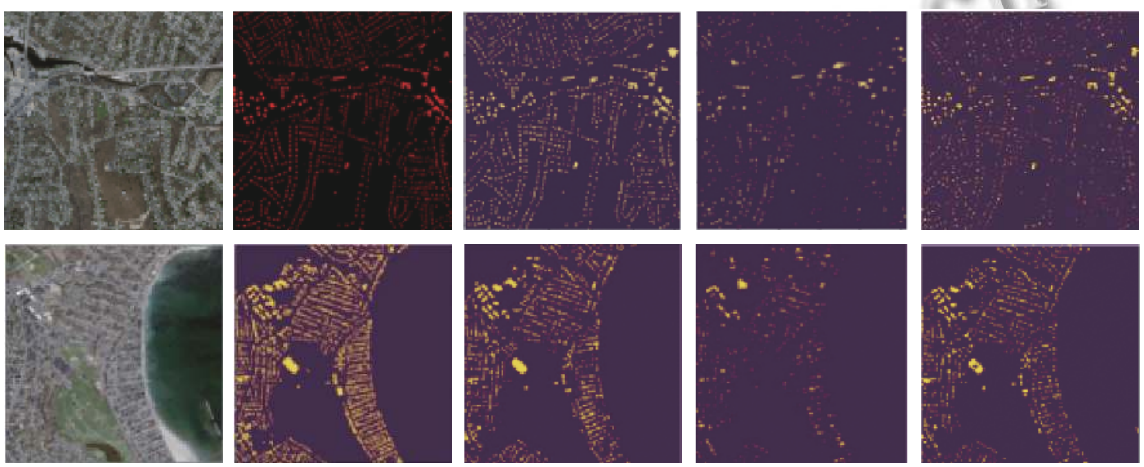
图 6 中对标注图像与改进 DeconvNet、DeconvNet 和 UNet 网络测试集上两组图像进行比较. 从图中可知改进 DeconvNet 建筑物提取比其他模型的效果要好, 充分说明改进 DeconvNet 能有效的对建筑物进行提取.

从表 3 为评估模型的指标, 在建筑物数据集上改进 DeconvNet 网络模型的综合指标要高于 DeconvNet 和 UNet 网络模型, 说明通过特征连接和数据的动态训练, 可以有效提高模型的分割精度.



(a) 像素精确度 (b) 损失函数

图 5 精确度与损失函数曲线图



(a) 原图 (b) 标注图像 (c) 改进 DeconvNet (d) DeconvNet (e) UNet

图 6 模型效果对比

4 结论

针对传统神经网络模型对遥感影像中小物体语义分割精度不足的问题, 本文采用编码与解码结构特征连接的方式改进反卷积网络, 有效地提升了小物体分割精度, 同时提出数据预训练模型, 对数据进行动态加载, 解决了网络模型过拟合问题. 在公开数据集上的实

验结果表明, 相比于 DeconvNet 和 UNet 模型, 改进 DeconvNet 模型能有效处理遥感影像的物体空间信息、边缘信息与像素之间的关系, 并有较好的像素准确率.

表 3 网络模型评估标准对比 (单位: %)

模型	*ACC*	*P*	*R*	*F*1
改进 DeconvNet	95	83	74	60
UNet	92	86	32	46
DeconvNet	89	57	12	20

参考文献

1 Martha TR, Kerle N, Van Westen CJ, *et al*. Segment optimization and data-driven thresholding for knowledge-based landslide detection by object-based image analysis. IEEE Transactions on Geoscience and Remote Sensing, 2011, 49(12): 4928–4943. [doi: 10.1109/tgrs.2011.2151866]

2 Yao XW, Han JW, Cheng G, *et al*. Semantic annotation of high-resolution satellite images via weakly supervised learning. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(6): 3660–3671. [doi: 10.1109/TGRS.2016.2523563]

3 李月臣, 杨华, 刘春霞, 等. 土地覆盖变化遥感检测方法. 水土保持研究, 2006, 13(1): 209–216. [doi: 10.3969/j.issn.1005-3409.2006.01.070]

4 潘朝. 多尺度显著性引导的高分辨率遥感影像建筑物提取. 科技创新与生产力, 2017, (5): 106–109. [doi: 10.3969/j.issn.1674-9146.2017.05.106]

5 黄亮, 宋晶. 一种改进迭代条件模型的遥感影像语义分割方法. 软件导刊, 2019, 18(1): 183–185.

6 李欣, 唐文莉, 杨博. 利用深度残差网络的高分遥感影像语义分割. 应用科学学报, 2019, 37(2): 282–290.

7 苏健民, 杨岚心, 景维鹏. 基于 U-Net 的高分辨率遥感图像语义分割方法. 计算机工程与应用, 2019, 55(7): 207–213. [doi: 10.3778/j.issn.1002-8331.1806-0024]

8 Noh H, Hong S, Han B. Learning deconvolution network for semantic segmentation. Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile. 2015. 1520–1528. [doi: 10.1109/ICCV.2015.178]

9 Zeiler MD, Taylor GW, Fergus R. Adaptive deconvolutional networks for mid and high level feature learning. Proceedings of 2011 International Conference on Computer Vision. Washington, WA, USA. 2012. 2018–2025. [doi: 10.1109/ICCV.2011.6126474]

10 Dumoulin V, Visin F. A guide to convolution arithmetic for deep learning. arXiv: 1603.07285, 2016.

11 Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich, Germany. 2015. 234–241. [doi: 10.1007/9783-319-24574-4_28]

12 Bjorck J, Gomes C, Selman B, *et al*. Understanding batch normalization. arXiv: 1806.02375, 2018.

13 Erb RJ. Introduction to backpropagation neural network computation. Pharmaceutical Research, 1993, 10(2): 165–170. [doi: 10.1023/A:1018966222807]

14 Booth DE. The cross-entropy method. Taylor & Francis Group, 2008. https://www.tandfonline.com/doi/citedby/10.1198/tech.2008.s534?scroll=top&needAccss=true.

15 Mnih V. Machine learning for aerial image labeling[Ph.D. Thesis]. Canada: University of Toronto, 2013.