

基于改进的无上下文网络解决拼图问题

潘佳俊*, 宗旭磊#, 杨世兴

北方工业大学理学院, 北京

收稿日期: 2024年11月27日; 录用日期: 2024年12月21日; 发布日期: 2024年12月31日

摘要

本文提出了利用深度学习技术解决拼图问题的新方法。拼图游戏作为一种跨学科的智力挑战, 其复杂性随着碎片数量的增加而指数级增长。本文介绍了一种基于改进AlexNet卷积神经网络架构的无监督学习方法, 用于解决自然图像的拼图问题。该方法通过构建一个拼图任务预测网络, 提高了模型的性能和泛化能力。实验部分在自制的果蔬数据集上进行, 验证了模型在不同复杂度拼图还原任务中的有效性。最后, 文章总结了研究成果, 并对未来的研究方向提出了展望。

关键词

图像恢复, AlexNet, 深度学习

Solving Jigsaw Puzzle Problem Based on Improved Context-Free Network

Jiajun Pan*, Xulei Zong#, Shixing Yang

College of Science, North China University of Technology, Beijing

Received: Nov. 27th, 2024; accepted: Dec. 21st, 2024; published: Dec. 31st, 2024

Abstract

In this paper, we propose new methods for solving jigsaw puzzles using deep learning techniques. As an interdisciplinary intellectual challenge, the complexity of jigsaw puzzles grows exponentially with the number of pieces. This paper introduces an unsupervised learning method based on a modified AlexNet convolutional neural network architecture for solving the jigsaw puzzle problem of natural images. This method improves the performance and generalization ability of the model by constructing a puzzle task prediction network. The experimental part is carried out on the self-

*第一作者。

#通讯作者。

made fruit and vegetable dataset, and the effectiveness of the model in puzzle restoration tasks of different complexity is verified. Finally, the paper summarizes the research results and puts forward the future research directions.

Keywords

Image Restoration, AlexNet, Deep Learning

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

拼图游戏是一款广受欢迎的智力游戏，它具有多样的形式，不同的难度，玩家需要将杂乱破碎的小碎块按照一定的规律还原成完整的原始图片。解决拼图游戏是一种十分复杂的问题，它涉及计算机科学，数学，工程学等学科。学者们利用这些学科知识将无序的小碎片按照某些规则还原成有序的原始图片，提出了许多拼图问题的相关研究，并将这些研究成果应用到实际生活中，例如辅助天气检测预测[1]、破碎的考古文物的修复[2]-[4]、中等教育中的应用[5]、基因生物学中 DNA 的组装[6]。规范的拼图游戏是将图片分割成无数的小碎片，其难度在于这些碎片的无序性、数量、形状以及图片本身的图像复杂度都会很大程度地决定拼图问题的复杂度。随着拼图块数的增加，拼图的复杂度也相应增加，导致计算量十分巨大。若拼图问题的规模为 $N \times N$ ，那么在方向和位置都不确定的情况下就会有 $N^2! \times 4^{N^2}$ 种可能的拼接结果，这是十分巨大的计算量。而且拼图块的形状和图案可能存在信息的缺失和模糊，这使得自动还原拼图系统在区分和放置特定的拼图块时出现错误，这是一个比较大的困难。

早在 1964 年，Freeman 和 Garder [7] 就针对均是灰色拼图块，仅依靠拼图块的形状来实现拼图还原展开了相关研究，这是历史上第一次提出的关于图像拼图问题的自动求解方法，这是一个开创性的工作。但这种仅依赖于拼图块的轮廓信息进行拼图还原的算法在后来的发展中逐渐显现弊端。当处理形状简单且相似的拼图块时，该方法不能清楚的区分它们。主要是由于它缺乏对图像内容的利用，忽略了图像丰富的细节。最终 Demaine 等人证明了无底层图像拼图、边缘匹配拼图和组合体拼图(polyomino packing)谜题都是 NP 完全的[8]。于是学者们同时利用拼图块图像的颜色和形状信息，对其进行混合，完成拼图的自动还原。Jin 等人[9]提出了既考虑拼图块之间的边缘相似性，又要衡量拼图块之间的内容相似性。并且提出一种算法，计算两拼图之间的内容相似性得分，并组合内容相似性得分和边缘相似性得分，最终找出最佳匹配的拼图块。Zhi 等人[10]针对在颜色匹配中，不同的图像可能具有相同的直方图这个问题，提出了一种结合边缘和颜色特征的算法。该算法首先进行边缘特征提取，然后选取合适的边缘区域计算直方图，采用直方图相交算法进行拼接。这种根据颜色和边缘进行匹配的方法，虽可以有效地解决拼图问题，但是由于噪声等因素的影响对输入的拼图块的质量要求较高。

随着深度学习技术的发展，相关技术也应用在解决拼图问题上。一些基于卷积神经网络的方法被大量引入。Le 等人提出在图像重组的局部阶段通过构建深度卷积网络(CNN)来检测两两拼图块之间的兼容性，并且正确地计算出成对的匹配。在全局组合阶段使用两种基于循环闭合的搜索算法。通过该两个阶段，可解决各种谜题[11]。但是这种方法在构建和训练深度卷积网络可能会需要大量的计算资源，而且当训练数据不充足时，卷积模型在处理未见过的图像类型时可能会表现不佳。目前深度学习在拼图还原算法中的应用日渐成熟，不仅在图像恢复领域有着重大的突破，还在拼图问题中展现出了巨大的潜力和灵

活性。随着研究的深入,未来可能出现更高效、便捷的方法,以解决更复杂的拼图还原问题。

关于方形拼图问题 Gallagher 等人将其分为三种类型[12]。我们将专注于其中一种类型——假设每个拼图块的方向是已知的,只有拼图块的位置是未知的,来解决拼图问题。这是传统意义上的谜题。

拼图问题本质上就是一个“排序”问题,目标为根据拼图块之间的空间关系进行排序,而卷积神经网络(CNN)特别适合处理空间数据。因此我们借鉴 Noroozi [13]等人提出的无监督学习方法,并在此方法的基础上对其进行改进,用来解决自然图像的拼图问题。我们通过对经典的 AlexNet 架构进行了针对性地改进,以适应特定的图像处理任务,基于该特征提取器搭建拼图任务预测网络,为拼图块排列顺序的预测问题提供了一种新的解决方案。这些改进和设计不仅提高了模型的性能,还增强了其泛化能力,为自动化拼图任务提供了一种有效的技术手段。

2. 基本网络架构与评价指标

2.1. AlexNet 卷积神经网络

AlexNet 卷积神经网络是由 Krizhevsky、Sutskever 和 Hinton 于 2012 年提出的一种经典卷积神经网络。彼时, AlexNet 在视觉识别领域斩获了卓越的成绩,使得 CNN 成为在图像分类上的核心算法模型。

AlexNet 由卷积层和全连接层构成,其中卷积层共计五层,每一层卷积层均包含卷积核、偏置项、ReLU 激活函数以及局部响应归一化(LRN)模块。全连接层共有三个。其基本配置如下:

卷积层 C1: 使用 96 个尺寸为 $11 \times 11 \times 3$ 、步长为 4 的卷积核对 RGB 三通道的 $224 \times 224 \times 3$ 输入图片进行滤波操作。接着,将一对 $55 \times 55 \times 48$ 的特征图分别输入 ReLU 激活函数。对激活后的图像实施最大池化。池化后进行 LRN 处理。卷积层 C2: 使用 C1 输出作为输入,使用 256 个大小为 $5 \times 5 \times 48$ 的卷积核对其进行滤波。卷积层 C3、C4、C5: 分别采用 3×3 的卷积核,步长为 1,滤波器数分别为 384、384 和 256。全连接层 F6、F7: 通过 4096 个大小为 $6 \times 6 \times 256$ 的卷积核生成 4096 个尺寸为 1×1 的特征图。将这些特征图输入 ReLU 函数并且进行 Dropout 处理,以减少过拟合现象的出现。全连接层 F8: 具备 1000 个神经元,对应着 1000 个类别的输出。最后这些输出输入到 1000 维的 softmax 层,生成 1000 个类别预测的值。

这种多卷积层和全连接层的深度结构使得 AlexNet 能够有效提取每个拼图块中的高层次特征,如形状、颜色和纹理,从而更好地处理拼图块之间的复杂组合关系,提高拼图重组的准确性。

2.2. 无上下文网络(CFN)

Context-Free Network (CFN)是 Noroozi 等人在 2016 年提出的孪生卷积网络[13],主要用于图像检索和图像匹配任务。该网络通过对输入的图像对进行处理,学习到两张图像的相似性,能够有效地提取图像的高层特征。它使用共享权重的 AlexNet 架构直到第一层全连接层(F6),并将所有 F6 层的输出连接起来作为 F7 的输入。CFN 在 ImageNet2012 数据集上的分类任务中表现与 AlexNet 相当,但参数更少,只有 2750 万个,而 AlexNet 有 6100 万个参数。CFN 的 F6 层参数约为 2M,而 AlexNet 的 F6 层参数约为 37.5M。CFN 的 F7 层参数比 AlexNet 多 2M。这种结构适用于不同的任务,如检测和分类。

2.3. 评价指标

对拼图碎片进行还原,相应的评价指标是必不可少的,通过计算指标的度量值,有利于拼图还原的调整与优化。

自动解决拼图游戏涉及很多相关的领域,其中 Alajlan N. [14]总结了相关拼图算法优劣的评判准则,具体如下:

Accuracy (准确性): 算法应该以高度的准确性组合拼图谜题。

Invariance (不变性): 一个好的拼图算法应该对旋转和平移拼图块等变化保持良好的不变性。

Robustness (鲁棒性): 当某些数据部分缺失、多余、重叠或存在大量噪声时, 算法应该仍能正常运行且表现良好。

Scalability (可伸缩性): 随着拼图问题规模数量的增加, 相应算法的时间复杂度和空间复杂度应该保持在一定范围不变。

Generality (通用性): 拼图算法可以应用于不同类型的拼图图像, 如黑白图, 灰度图和彩色图像。

Computational complexity (计算复杂性): 为了适合实时应用程序, 拼图算法应该具有较良好的计算效率。

3. 具体模型与数据介绍

3.1. AlexNet 特征提取器

本部分网络结构借鉴了经典的 AlexNet 架构, 并针对特定任务的需求进行了系统调整。首先, 输入层接收三通道的 RGB 图像, 默认图像尺寸为(64 × 64)。接着是卷积层、池化层与激活函数部分。在每个卷积层中均使用了 Conv2D, 且每层的卷积核数量不同(例如 96、256 等), 以提取出多层次的特征。每层卷积层之后均加入了批量归一化(Batch Normalization)处理, 这有助于加速训练过程并稳定模型。激活函数使用 SELU (Scaled Exponential Linear Unit), 其公式如下:

$$\text{SELU}_x = \lambda \begin{cases} x, & \text{if } x > 0 \\ e^x - 1, & \text{if } x \leq 0 \end{cases}$$

其中 λ 是一个缩放参数, 通常取值为 $\lambda = 1.0507$, 此参数旨在确保当输入数据的均值为 0, 方差为 1 时, 经过 SELU 激活函数后输出的均值接近于 0, 方差接近于 1, 这种设计使得 SELU 在正输入区域保持线性, 而在负输入区域提供指数增长, 有助于网络捕捉复杂的非线性关系。该激活函数特别适合深度神经网络, 相比 ReLU 激活函数, SELU 对噪声更具有鲁棒性, 使其在处理含有噪声的数据时表现的更好, 并且由于其自归一化的特性, SELU 也有助于网络训练的稳定性和收敛速度。MaxPooling2D 池化层在减小数据尺寸的同时保留关键信息, 从而提高模型的计算效率。具体而言, 各层的参数如下: 第一层卷积有 96 个大小为(11, 11)、步长为(2, 2)的卷积核, 采用“same”填充方式, 这使得输出尺寸保持不变; 第二层卷积使用 256 个(5, 5)的卷积核, 填充方式为“valid”(即无填充), 这会减少输出尺寸; 第三层卷积配备 384 个(3,3)的卷积核, 采用“same”填充; 第四层同第三层结构; 第五层卷积则是 256 个(3, 3)的卷积核, 亦采用“same”填充。随后是扁平化层, 将三维特征图展平为一维, 以便后续传入全连接层。最后看全连接层, 这是一个拥有 1024 个神经元的全连接层, 通过 SELU 激活函数实现非线性化, 并且应用了正则化来减少过拟合现象。网络架构示意图如图 1。

3.2. CFN 拼图任务预测网络

此模型结构使用多个 AlexNet 模型作为输入, 以预测拼图块的排列顺序。具体结构如下: 首先, 定义 n^2 个输入层(每一个输入层代表一个拼图块的输入), 将这 n^2 个输入分别传递给 AlexNet 特征提取器, 以特征提取, 得到 n^2 个输出特征图。然后对这些特征图使用 Concatenate 进行拼接, 以组合特征信息。最后, 将重组后的特征图送入全连接层中。全连接层共有两层, 第一层为 2048 个神经元的全连接层, 通过 SELU 激活函数激活, 并应用正则化防止过拟合。第二层输出排列(这里为 $n^2!$), 使用 Softmax 激活, 给出 $n^2!$ 对种拼图排列方式的概率。

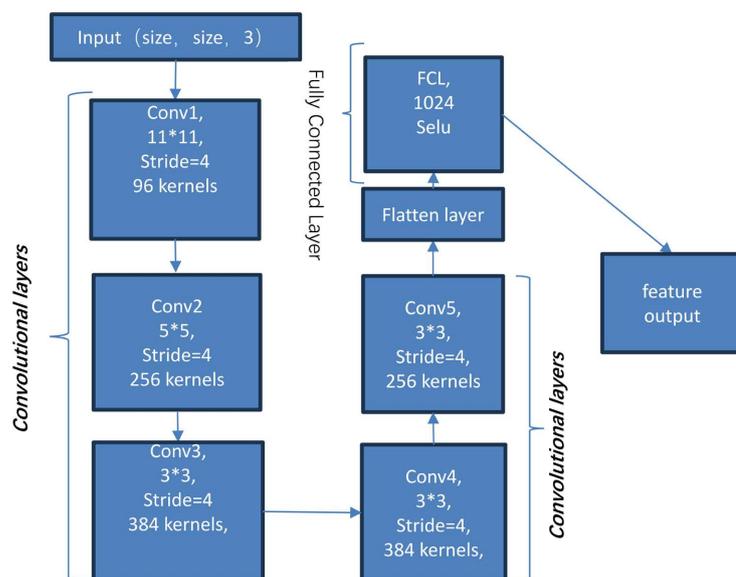


Figure 1. Schematic diagram of the AlexNet feature extractor architecture

图 1. AlexNet 特征提取器架构示意图

3.3. 数据介绍

本数据集包含约 15,000 张图片，均为个人网络收集和制作。这些图片涵盖了日常生活中常见的果蔬种类，如苹果、粳粳干、白萝卜、菠萝、车厘子等，共计 81 类，这样的分类有助于模型学习不同果蔬之间的视觉差异，提高识别的准确性。为了提高模型的泛化能力和减少过拟合，我们采用数据增强技术来增加数据量。分别对图片进行了如下操作：

- 1) 水平反转，帮助模型学习到图像的对称性。
- 2) 进行不同角度的旋转，模拟果蔬在实际环境中可能出现的不同朝向。
- 3) 对图片进行缩放，以模拟不同距离下的视觉效果。
- 4) 调整图片的亮度、对比度、饱和度，以模拟不同光照条件下的视觉效果。

这是一个相对较大的样本量，有助于训练出一个泛化能力强的模型。

我们按照 4:1 的比例将数据集划分为训练集和测试集。这意味着大约 12,000 张图片将用于训练，而 3000 张图片将用于测试。这样的划分比例有助于模型在训练过程中学习到足够的特征，同时在测试集上评估模型的性能。

在开始训练之前，需要对数据进行预处理，以确保数据的质量和一致性。预处理步骤如下：

首先，将所有图片调整为统一的尺寸，例如 224×224 大小的图像，以适应网络的输入要求。然后，检查图像是否存在模糊、过曝和欠曝的问题。最后，为每一类别分配一个唯一的标签，并确保标签的正确性。

4. 实验结果与分析

4.1. 实验设置

本实验在自制的果蔬数据集上展开，旨在测试所提出模型的性能。实验环境配置为 AMD Ryzen 9 6900HX with Radeon Graphics 3.30 GHz 处理器、40.0GB 内存(39.3GB 可用) (RAM)，并基于 TensorFlow 深度学习框架进行模型的训练和评估。所有实验均在相同条件下进行，以确保结果的一致性和可重复性。

4.2. 结果分析

经过实验测试, CFN 拼接任务预测网络模型在较低复杂度的拼图还原任务(2×2 拼图)上取得了一定的成果, 随着数据量的增加, 拼图还原的准确率也随之增加, 如表 1 所示。

Table 1. Comparison of accuracy for various data

表 1. 各数据准确率对比

总数据量	正确数据量	模型预测准确率
99	69	69.70%
474	348	73.42%
906	668	73.73%
1386	1062	76.62%
2157	1655	76.73%
4203	3178	75.61%

另外我们也与现有技术相比, 本模型在准确性方面表现出一定优势, 能够较为正确地恢复拼图, 复原后的图像如图 2 所示。与传统的特征提取方法如 SIFT 和 HOG 相比, 本模型无需依赖手工设计的特征, 而是能够自动从数据中学习特征, 因此其效果不易受到特征选择的影响。尽管更复杂的深度学习算法如 VGG 和 ResNet 在性能上表现更为优越, 但它们通常结构复杂, 计算资源消耗较大。而本模型则具有更为简洁的结构, 能够根据不同任务的需求进行灵活调整, 并且在小规模数据集上表现出色, 效率高, 能够有效平衡性能与效率之间的关系。这表明, 本模型在不同规模的拼图问题中具有较强的适应性。与结构相对简单的 LeNet 模型相比, 实验结果显示本模型在特征提取和复杂图像识别能力上更具优势, 适合处理更复杂的拼图问题。在处理复杂果蔬图像时, 能够更精准地还原拼图碎片的位置, 减少错误拼接。例如, 在处理具有相似颜色和形状的水果拼图时, 如苹果和橙子部分区域颜色相近, 但模型仍能准确区分并放置拼图块。

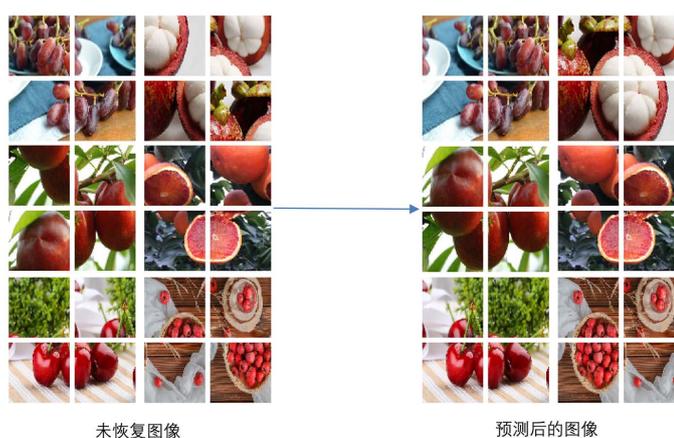


Figure 2. Before and after puzzle image
图 2. 拼图前后图像

在不变性方面, 模型对图像的平移和旋转具有较好的适应性。在数据增强过程中, 通过对图片进行

不同角度旋转和水平反转操作, 模型学习到了图像在不同变换下的特征, 从而在实际拼图还原中, 即使图像存在一定程度的旋转或平移, 也能准确识别拼图块的位置。

鲁棒性方面, 模型在面对一定程度的数据噪声和图像部分缺失时, 仍能保持较好的性能。尽管在数据预处理中已尽量避免图像质量问题, 但实际数据中仍可能存在少量噪声, 模型能够在这种情况下完成拼图还原任务, 不过当噪声程度超过一定阈值时, 性能会有所下降。

在可伸缩性上, 随着拼图问题规模的增加, 模型的计算复杂度在可接受范围内。但当拼图规模增大到一定程度, 计算时间会有所增加。例如, 对于较大尺寸的拼图图像(如 5×5 或更大), 模型的处理时间相较于小尺寸拼图(如 2×2 或 3×3)明显增长。

通用性方面, 模型在处理不同类型的果蔬图像时表现良好, 无论是形状规则的水果(如苹果、橙子)还是形状不规则的蔬菜(如西兰花、胡萝卜), 均能有效还原拼图。

然而, 模型也存在一定局限性。计算复杂性方面, 尽管在当前实验条件下表现尚可, 但对于大规模实时拼图应用场景, 仍有优化空间。例如, 在处理高清或超大尺寸拼图时, 计算速度可能无法满足实时性要求。数据依赖性方面, 模型的性能高度依赖于训练数据的质量和多样性。若训练数据中某些果蔬种类的图像数量不足或特征不明显, 可能导致模型在处理此类图像拼图时准确率下降。

5. 总结与展望

本研究提出的基于改进的 AlexNet 特征提取器和 CFN 拼图任务预测网络相结合的方法, 成功实现了对果蔬图像的拼图块的准确预测。实验结果显示, 模型在准确率和其他评估标准上表现良好, 验证了模型的有效性。未来研究可考虑进一步优化网络结构, 提升计算效率, 同时增加训练数据的丰富度和多样性, 以提高模型的性能和泛化能力, 拓展其在更广泛拼图问题中的应用。

基金项目

北京市大学生创新创业训练项目(10805136024XN139-182)。

参考文献

- [1] 陈鲍发, 马中元, 徐芬, 等. 天气雷达风暴跟踪信息拼图技术设计与应用[J]. 气象科学, 2020, 40(6): 838-848.
- [2] 周明全, 税午阳, 武仲科, 等. 交互式破碎文物虚拟修复方法[P]. 中国专利, CN201410006146.7. 2024-11-13.
- [3] 王克刚. 基于学习优化与信息融合的陶瓷文物碎片分类研究[D]: [博士学位论文]. 西安: 西北大学, 2017.
- [4] 袁洁, 周明全, 耿国华, 张雨禾. 基于 Morse-Smale 拓扑特征的文物碎片拼接算法[J]. 自动化学报, 2018, 44(8): 1486-1495.
- [5] 胡继飞. 拼图式合作学习及其在教师培训中的应用[J]. 外国中小学教育, 2009(6): 43-46+57.
- [6] Marande, W. and Burger, G. (2007) Mitochondrial DNA as a Genomic Jigsaw Puzzle. *Science*, **318**, 415-415. <https://doi.org/10.1126/science.1148033>
- [7] Freeman, H. and Garder, L. (1964) Apictorial Jigsaw Puzzles: The Computer Solution of a Problem in Pattern Recognition. *IEEE Transactions on Electronic Computers*, **13**, 118-127. <https://doi.org/10.1109/pgec.1964.263781>
- [8] Demaine, E.D. and Demaine, M.L. (2007) Jigsaw Puzzles, Edge Matching, and Polyomino Packing: Connections and Complexity. *Graphs and Combinatorics*, **23**, 195-208. <https://doi.org/10.1007/s00373-007-0713-4>
- [9] Jin, S.-Y., Lee, S., Azis, N.A. and Choi, H.-J. (2014) Jigsaw Puzzle Image Retrieval via Pairwise Compatibility Measurement. 2014 *International Conference on Big Data and Smart Computing (BIGCOMP)*, Bangkok, 15-17 January 2014, 123-127. <https://doi.org/10.1109/bigcomp.2014.6741421>
- [10] Zhi, L., Ge, Q. and Ji, Z. (2009) Image Matching Algorithm Based on Edge Color Used in Automatic Computer Jigsaw Puzzle. 2009 *2nd International Symposium on Knowledge Acquisition and Modeling*, Wuhan, 30 November-1 December 2009, 269-272. <https://doi.org/10.1109/kam.2009.149>
- [11] Le, C. and Li, X. (2019) Jigsawnet: Shredded Image Reassembly Using Convolutional Neural Network and Loop-Based

-
- Composition. *IEEE Transactions on Image Processing*, **28**, 4000-4015. <https://doi.org/10.1109/tip.2019.2903298>
- [12] Gallagher, A.C. (2012) Jigsaw Puzzles with Pieces of Unknown Orientation. 2012 *IEEE Conference on Computer Vision and Pattern Recognition*, Providence, 16-21 June 2012, 382-389. <https://doi.org/10.1109/cvpr.2012.6247699>
- [13] Noroozi, M. and Favaro, P. (2016) Unsupervised Learning of Visual Representations by Solving Jigsaw Puzzles. In: Leibe, B., Matas, J., Sebe, N. and Welling, M., Eds., *Computer Vision ECCV 2016*, Springer, 69-84.
- [14] Alajlan, N. (2009) Solving Square Jigsaw Puzzles Using Dynamic Programming and the Hungarian Procedure. *American Journal of Applied Sciences*, **6**, 1941-1947. <https://doi.org/10.3844/ajassp.2009.1941.1947>