《智能信息处理》课程考试

基于概念格语义分析的图像标注

覃文强

考核	到课[10]	作业[20]	考试[70]	课程成绩[100]
得分				

2021 年 11 月 22 日

基于概念格语义分析的图像标注

覃文强

（大连海事大学 信息科学技术学院, 大连 116026）

摘　要　随着互联网中图像资源的爆炸式增长，如何使用语义标签快速进行图像检索成为亟待解决的问题之一。传统的图像标注工作大多由人工完成，存在主观性和随意性，容易造成图像标签缺失、标注错误的现象。因此，如何有效地进行图像完备标注、丰富图像语义内容成为一个重要的研究课题。CNN（卷积神经网络）因摒弃了复杂的特征融合过程，能够通过自主学习图像特征，更好地获取图像视觉信息。形式概念分析是一种有效的语义层次分析方法。为有效完善图像语义标签，本文基于卷积神经元网络和形式概念分析，对图像的语义完备自动标注进行研究。在提出的基于 CNN 与概念格语义分析的图像完备标注算法中，首先利用卷积神经网络进行图像视觉特征的学习，结合概念格进行图像标签语义分析，最后对候选标签集的标签进行排序，实现最终标注。

关键词　形式概念分析；概念格；图像标注

Image Annotation based on semantic Analysis of concept Lattice

Qin Wenqiang

(School of Information Science and Technology, Dalian Maritime University, Dalian 116026)

Abstract　With the explosive growth of image resources in the Internet, how to use semantic tags for image retrieval has become one of the urgent problems to be solved. Most of the traditional image labeling work is done manually, which has subjectivity and arbitrariness, which is easy to cause the phenomenon of missing image label and wrong labeling. Therefore, how to effectively complete image annotation and enrich image semantic content has become an important research topic. Because of abandoning the complex feature fusion process, CNN (convolutional neural network) can obtain better image visual information by learning image features independently. Formal concept analysis is an effective method of semantic analytic hierarchy process. In order to effectively improve the image semantic tagging, this paper studies the semantic complete automatic tagging of images based on convolution neural network and formal concept analysis. In the proposed image complete annotation algorithm based on CNN and concept lattice semantic analysis, firstly, the convolution neural network is used to learn the image visual features, and then the concept lattice is used to analyze the image label semantics. Finally, the tags of the candidate label sets are sorted to achieve the final annotation.

Key words　Formal concept analysis; Formal background; Image annotation

1　图像标注

随着社会媒体和大数据的快速发展，用户上传和共享了数以万亿的数字图像，越来越多的图像被用户进行了不同程度的初始标记。图像标签标注是给予图像标记的一种有效方式，网络图像数据的爆炸性增长以及图像标记的主观性和随意性，造成了大量图像的标签缺失和语义噪声，不能很好地描述图像内容，为丰富图像标签内容，提高图像检索准确率，许多研究者已对缺失标签进行自动补全的图像标签完备方法展开了深入研究[1,2]；同时，大部分的自动标注算法假设训练集标注的标签是完备

的，但实际上与现实世界图像数据表达的内容相差甚远，也表现出图像标注标签的不完备性。

深度卷积神经网络因具有深层网络结构、能够主动学习并抽象出图像的视觉特征，具有更强大的表达能力，在各种视觉识别任务中显示出巨大潜力。形式概念分析因能很好地将概念之中包含与被包含、上下层次之间的关系展示出来，成为一种高效数据分析与知识提取的工具。建立缺失标签图像低层的视觉特征与标签之间的关系，并有效地改善语义标签是提高图像标签完备标注精度的一个关键。因此，基于深度卷积神经网络和形式概念分析，研究一种图像非完备自动标注方法以提高图像语义检索效率的，具有重要的理论和应用研究价值。

2 形式概念分析

2.1 形式概念分析方法

形式概念分析(Formal concept Analysis，简称FCA)是一种有效的数据分析方法和知识表示工具。形式概念分析理论最早由德国数学家 Wille 教授在 1982 年提出。Ganter 教授在 1999 年的学术著作中概括并总结了早期的形式概念知识框架及理论成果。Wille 教授将 “形式概念”中的定义进行了详细的数学描述。在这种数学描述中“父概念”，“子概念”的关系将是偏序关系，按这个关系，任何一个概念集合均有上确界和下确界，因而按照这个关系，全部概念集合将是一个“完全格”。由于有“格理论”强大的支持，所以近年来形式概念分析理论得到了比较广泛的应用。

2.2 概念格的定义

格（lattice）的意义是任两个元素的上确界和下确界都存在的偏序集。完备格为任一子集的上确界和下确界存在的偏序集，其特点是只有一个最高点，且只有一个最低点，且图中任何两点连通。

概念格(Concept Lattice)是 20 世纪 80 年代初由德国 Wille 教授[3]提出的数据分析工具。概念格作为形式概念分析中的一种重要的格结构，描述了概念节点外延和内涵间的本质联系。概念格是元素为概念的完备格，概念格的每个节点是一个形式概念，每一个形式概念都是由外延和内涵两部分组成。概念格通过 Hasse 图生动和简洁地体现了这些概念之间的泛化和特化关系。从形式背景中生成概念格的过程实质上是一种概念聚类过程。因为能很好地展示概念中包含与被包含、上-下层次之间的关系以及对象与属性之间的所属关系，所以能达到高效的数据分析与知识提取的目的。参照文献[4]，本文给出相关定义：

定义 1. 在概念格理论之中，一般会将形式背景作为一个三元组 C(U , A , R)，在这之中，对象集即为 U，属性集即为 A，及一个二元关系。若对于一个对象与任意属性，存在关系 R，那么称为“对象 u 具有属性 a”，记为 uRa。如表 1 所示，表中用“x”标记出对象与属性之间的映射关系。

表 2.1 对象 U 与属性 A 的形式背景

对象集 U	属性集 A						
	a_1	a_2	a_3	a_4	a_5	…	a_n
u_0	x	x		x	x	…	x
u_1	x				x	…	
u_2						…	
u_3	x		x	x		…	x

定义 2. 对于任意一个二元组 z =(I,T)，I⊆U ，T⊆A，在对象集和属性集上分别满足如下运算：

$$f(I)=\{a\in A\mid \forall u\in I, uRa\}$$

$$g(T)=\{u\in U\mid \forall a\in T, uRa\}$$

若 f(I) =T ，g(T) =I ，则定义 z =(I,T)是基于形式背景 C(U , A ,R) 这一基础之上的形式概念，所以形式概念 z 的外延即为 I，而形式概念 z 的内涵即为 T。

定义 3. 设$z_1=(I_1,T_1)$、$z_2=(I_2,T_2)$表示形式背景 C(U , A , R) 上的两个形式概念，若

$$z_1\le z_2\leftrightarrow I_1\subseteq I_2\leftrightarrow (T_1\subseteq T_2)$$

则z_1是z_2的子类节点，z_2是z_1父类节点。将用这种偏序关系组成的集合称为 C 上的概念格，记为<L(U, A, R),≤>，其中≤表示概念格内节点之间的偏序关系，同时，根据形式背景 C 中的偏序关系可以得到相应概念格的 Hasse 图。

2.3 概念格的构建

概念格的表示形式是 Hasse 图，概念格的构建的基础是形式背景，形式背景描述了多个形式概念之间的关系，单个形式概念描述了形式对象以及形式对象所具有的形式属性之间的关系。所以，概念格的构建必须明确不同形式对象以及不同形式对象所具有的形式属性。

概念格的构建包含以下几个步骤：生成形式背

景，约简形式背景，生成单值形式背景，确定父子关系，绘制 Hasse 图，补充各形式概念的上确界和下确界，最后获得概念格。

3 基于CNN和概念格的图像完备标注

基于深度卷积神经网络的图像自动标注方法研究已经得到了越来越多的关注，并取得了很多显著的成果。但仍然存在一些问题需要进一步改善与提高。首先，卷积神经网络需要大量被标注的图 像样本进行训练。目前存在开源的大型图像数据集如ImageNet[5]，可以很好的满足模型对样本数据量的需求，但对于样本数量较少的图像数据集，不能很好的完成模型训练；第二，图像数据集中通常会存在一些样本相似度较高的类别，存在样本错误分类的问题；第三，数据集中不同类别的物体可能存在依存关系，例如天空和白云，如果在语义上不加处理，会影响图像的标注效果。针对上述问题，本章提出一种基于 CNN 和概念格语义扩展的图像完备标注模型。该模型利用 CNN 自动抽象图像特征的强大表达能力，以及概念格高效展示包含与层次关系的能力，对标签贡献值进行排序完成标签的预测，改善深度卷积神经网络的标注结果。

3.1 算法流程

在基于CNN和概念格的图像完备标注模型中，首先将带有初始噪声标签的待完备图像输入卷积神经网络，得到深度卷积特征图；其次用 softmax 分类器对待完备图像进行分类标注，作为初始标签集合 W，并将相似的标签放入近邻标签集 T 中；第三，获取相似图像，构成近邻图像集 I；第四，通过概念格语义计算近邻标签集 T 和近邻图像集 I 的相关度，得到候选标签集合 W ；最终，获得图像的完备标签。算法流程如图 3.1 所示。

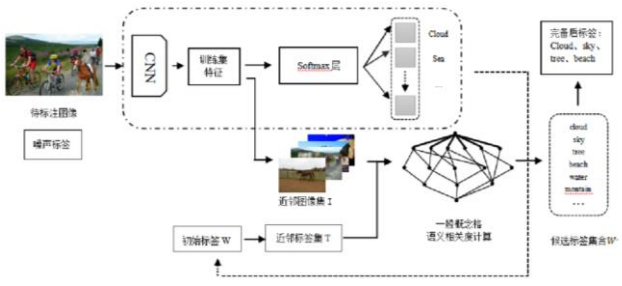

图 3.1 基于 CNN 和概念格语义扩展的图像标注流程图

3.2 基于 CNN 的图像初始标注获取

本文采用 VGG19 模型获取图像的初始标注，获取初始标注的简单流程如图 3.2 所示。

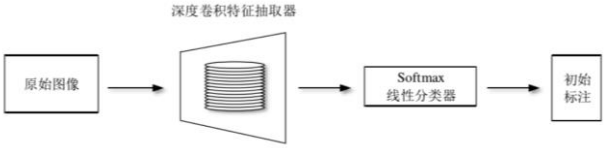

图 3.2 获取初始标注

具体步骤如下：

(1) 采用 VGG19 模型进行预训练，使用 ImageNet 数据集训练，并调试网络参数至最佳通用状态。

(2) 使用 Multi-Scale 做数据增强，将图像缩放到不同尺寸 S，满足输入要求。令处理后的待标注图像为 I_0，则 $I=[f_0,f_1,\ldots,f_m]^T$，其中 f_m 为原始像素，m 为像素数。

(3) 初始化图像标签数量，为得到图像的语义标签，为不陷入局部最优，减少过拟合，将其作为深度网络有效的监督信息，首先对图像进行 Normalized cut 分割，目的在于得到更多的图像块信息，为不陷入局部最优，减少过拟合状态，将其作为深度网络有效的监督信息进行训练，即

$$N\_cluster=n*inti\_labels \tag{3.1}$$

其中，N_cluster 为扩大后的聚类数，inti_labels 为初始标签数量，之后再在图像块中运用选择性搜索算法得到多个候选区域，最后对每一个候选区域进行标注，合并重复标签，得到最终的标注结果。

(4) 输入 CNN 网络，将高维的输入图像转化为低维的抽象的信号特征，将边缘特征抽象组合成更为简单的特征输出。

(5) 为减少卷积操作后存在的冗余信息及降低特征维数，采用最大池化操作。设第 i 层为池化层，输入的图像值为 f_i，分割成的图像块区域为 $R_k(k=1,2,\ldots,k)$，如式(3.2)：

$$pool(R_k)=\max_{i\in R_k} f_i \tag{3.2}$$

(6) 进行全连接层计算。对倒数第二个全连接层输出的 4096*1 的向量做 softmax 回归，得到特征向量，在得到的 20 个由深度网络提取到的特征做 softmax 回归得到标签的概率中选择最大的一个作为图像块的标签，重复该步骤直至所有图像块被标记，得到初始标注集合 W0，如图 3.3 所示。

然后，在获取图像初始标注数据之后，提出了

基于概念格进行语义分析的方法，进一步获取关联度较高的图像语义之间的特殊特征，用以提高标签标注精度。

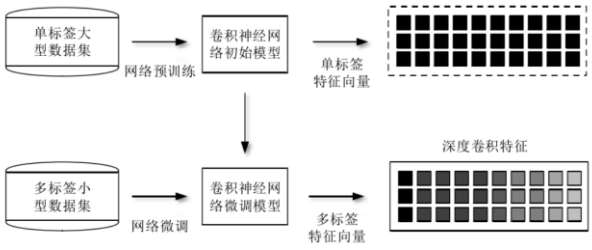

图 3.3 深度卷积特征获取

3.3 基于概念格的语义扩展改善

虽然 CNN 在图像标注领域取得了很大进展，能够逐层抽象特征图的重要信息，但由于方法本身高度依赖于真实边界框，当将其转移到没有任何边界框信息的多标签数据集时，可能会限制其泛化能力。如某幅图像存在缺失标签“cloud”” 和“sky”，经过卷积神经网络标注之后，只标注“cloud”，缺失了标签“sky”。但在一般情况下，“cloud”和“sky”并不完全孤立，存在依存关系，被用来标注同一幅图像的概率极高，如果在语义上不加处理，会影响图像的标注效果。本节将 VGG19 模型的 softmax 层去掉，首先，得到图像的 4096 维特征向量，并使用 PCA 进行降维操作保持 80%的特征差异，作为图像的最终特征；其次将得到的图像特征归一化之后转换为 256*256 的向量，若选择性搜索后得到的图像个数为 N，则图像转换成大小为 65535*N 的矩阵；最后对向量矩阵进行 SVD 分解，得到降序排列的特征值，利用特征值计算权值w_i，得到相似图像的权重，如式(3.3)所示：

$$w_i = \lambda_i \Big/ \sum_{i=1}^{n} \lambda_i \qquad (3.3)$$

其中，λ_i表示图像的特征值。将由深度网络得到的权重值w_i大于 0.5 的图像对应图像构成近邻图像集合 I，把近邻图像集与对应标签生成形式背景，利用图像标签之间的语义相关度来描述图像之间的相似程度，据此计算近邻图像的语义相关度。假设待标注图像I_0，得到 k 张(假设 k=5)与其最相似的近邻图像I_1-I_5，获取图像 I0 及其近邻图像所有的标签并入集合 T 中，则 I ={I_0，I_1,I_2，I_3,I_4,I_5}，假设 T={“sky”、“grass”、“river”、“tree”、“ground”、“people”、“bird”、“animal”、“dog”、“car”}。根据定义 1 构造近邻图像与标签映射关系并进行归一化处理，即存在映射关系“x”的将其置换为 1，反之，记为 0，构造出形式背景 G，如表 3.1 所示。为方便表示，分别用“t_1-t_{10}”按序表示标签集合中的词，并依形式背景 G 构造 Hasse 图，如图 3.4 所示。

表 3.1 形式背景 G 表

近邻图像	标签集合									
	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
I_0	1	1	0	1	1	0	1	1	0	0
I_1	1	0	0	0	1	0	0	1	0	0
I_2	0	0	0	0	0	0	0	1	0	1
I_3	1	0	1	1	0	0	0	0	0	0
I_4	0	1	0	1	0	1	0	1	1	1
I_5	0	0	1	0	0	1	1	1	1	0

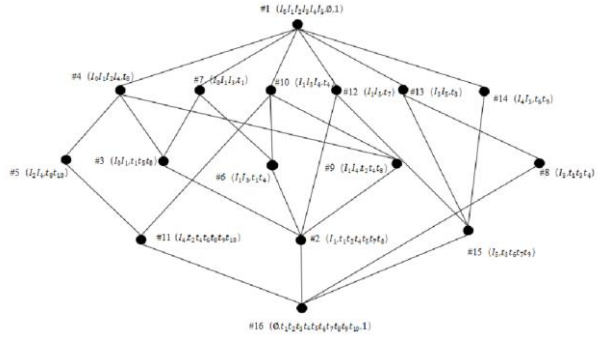

图 3.4 形式背景 G 的 Hasse 图

概念格是一种有效的语义层次分析工具，为利用概念格进行图像标签语义相关性分析，本文定义了如下一些语义相关度概念。

定义 5. 概念-概念相关度$\mathrm{Rel}(\mathrm{dist}(z_i, z_j))$。由图 3.6 可知，两个概念之间形成的通路越短，则概念间的相似度越大，若$\mathrm{Dist}(z_i, z_j)$表示一个格结构中两个概念之间形成通路的最短路径长度，则基于概念-概念之间的相关度计算公式如(3.4)定义如下：

$$\mathrm{Rel}(\mathrm{dist}(z_i, z_j)) = \tau/(\tau + \mathrm{Dist}(z_i, z_j)) \qquad (3.4)$$

其中，$\mathrm{Rel}(\mathrm{dist}(z_i, z_j))$表示概念$z_i$和概念$z_j$间的语义相关度；$\tau$为大于 0 的实数，这里取$\tau$=1。

定义 6. 外延-概念相关度 Rel(I, z)。随着深度的增加，由定义 2 可知，概念节点中外延数逐渐减少，共同拥有的内涵数就会越具体，概念之间的相似度也会随之减小。因此，本文通过考虑概念节点间的关系和概念节点所处的深度对图像语义相关度的影响，给出基于外延-概念的相关度计算公式如式(3.5)定义如下：

$$\mathrm{Rel}(\mathrm{I},\mathrm{z}) = (|\mathrm{I_i}| \cap |\mathrm{I_j}|)/\max((|\mathrm{I_i}|,|\mathrm{I_j}|)) \times (1+\sigma)^{(\mathrm{dep_1}+\mathrm{dep_2})} \quad (3.5)$$

其中，$\mathrm{Rel}(\mathrm{I},\mathrm{z})$表示的是基于外延-概念的相关度，$|\mathrm{I_i}| \cap |\mathrm{I_j}|$表示的是概念$\mathrm{z_i} = (\mathrm{I_i},\mathrm{T_i})$和概念$\mathrm{z_j} = (\mathrm{I_j},\mathrm{T_j})$间相同的外延个数；$\mathrm{dep_1}$和$\mathrm{dep_2}$分别代表的是概念节点$\mathrm{z_i}$和概念节点$\mathrm{z_j}$所处的深度，设概念格顶层概念的层次为 1，其节点深度为上邻节点概念层数加 1；σ是为体现概念节点深度对其影响的修正参数，这里取σ=0.1。

定义 7. 内涵-概念相关度$\mathrm{Rel}(\mathrm{T},\mathrm{z})$。概念格结构中，概念与概念之间距离越远，则外延所共同拥有的内涵数越少。由此可以得出，随着概念格 Hasse 图概念节点的深度增大，概念外延的语义相关度与外延共同拥有的内涵数成正相关性。因此，本文通过考虑概念节点间的关系和概念节点所处的深度对相关度的影响，提出基于内涵-概念的相关度计算公式如式(3.6)定义如下：

$$\mathrm{Rel}(\mathrm{T},\mathrm{z}) = (|\mathrm{T_i}| \cap |\mathrm{T_j}|)/\max((|\mathrm{T_i}|,|\mathrm{T_j}|)) \times (1+\sigma)^{(\mathrm{dep_1}+\mathrm{dep_2})} \quad (3.6)$$

其中，$\mathrm{Rel}(\mathrm{T},\mathrm{z})$表示的是概念-内涵的相关度，$|\mathrm{T_i}| \cap |\mathrm{T_j}|$表示的是概念节点$\mathrm{z_i}$和概念节点$\mathrm{z_j}$所拥有共同内涵数的个数；$\sigma$是修正参数，作用同定义 6。根据定义 5、定义 6 及定义 7，依据式(3.4)、式(3.5)、式(3.6)计算出每个概念节点之间的相关度$\mathrm{Rel}(\mathrm{z_i},\mathrm{z_j})$，降序排列得到近邻图像对其图像语义的支持度并将其归一化，利用相似图像之间的语义相关度，进一步衡量图像之间相似程度，可以大大减少噪声图像标签的加入。因此，综合考虑概念-概念、外延-概念、内涵-概念以上三者对图像语义相关度的影响，本文给出基于概念格的图像语义相关度公式(3.7)定义如下：

$$\mathrm{Rel}(\mathrm{z_i},\mathrm{z_j}) = \mathrm{Rel}(\mathrm{I},\mathrm{z}) \times \alpha + \mathrm{Rel}(\mathrm{T},\mathrm{z}) \times \beta + \mathrm{Rel}(\mathrm{dist}(\mathrm{z_i},\mathrm{z_j})) \times \gamma \quad (3.7)$$

其中，α、β、γ是各部分所占的权重比，且α + β + γ =1。由于内涵和外延在概念对中具有同等大小的权重比，根据概念格的对偶原则，本文取α = β =0.25，则γ =0.5。据此计算所有概念之间的语义相关度，如在形式背景 G 中，从节点#2 和#3、#3 和#4 存在上下位关系，节点#2 和#15 为同层次概念，由式(8)可以得出$\mathrm{Rel}(\mathrm{z_2},\mathrm{z_{15}}) < \mathrm{Rel}(\mathrm{z_3},\mathrm{z_4}) < \mathrm{Rel}(\mathrm{z_2},\mathrm{z_3})$。

由此可知，父节点的语义相似度要比同层次概念节点的高，同时，随着概念格层次的逐渐加深，父子节点之间的语义相似度也会随之增大。我们将包含同一对象的不同概念节点相关度叠加得到图像之间的语义相关度，例如由节点#2、#3、#4 可知待标注图像$\mathrm{I_0}$与图像$\mathrm{I_1}$的语义相关度为 1.322，与图像$\mathrm{I_5}$的语义相关度为 0.257。由此可得，待标注图像$\mathrm{I_i}$与训练集$\mathrm{I_j}$视觉相似度。当待标注图像$\mathrm{I_i}$越高时，图像$\mathrm{I_j}$与$\mathrm{I_i}$的语义相关度越高时，其标签贡献值越大，越有可能被标记。

3.4 标签预测

通过计算图像标签之间的语义相关度，获取一系列同待标注图像关联密切的近邻图像标签作为候选标签，对初始预测标签进行语义扩展。由于近邻图像与待标注图像的相似度程度不同，且一般与待标注图像语义相关度更相近的图像对标注结果影响更大。由于图像集 I 是根据图像底层特征搜索降序而得，并且同时考虑了底层特征与高层语义的相似性，兼顾近邻图像语义对标注结果的影响，从而避免某些标签过少或过多，改善标注结果，丰富图像的语义内容。因此，本文融合 CNN 标注结果并结合近邻图像与待标注图像的语义相关度，从视觉和语义两个角度，筛选候选标签集中关联程度强的候选标签，从而保留支持度更高的标签标记图像。根据式获取的图像块权重大小w_i，从视觉角度，将其作为近邻图像$\mathrm{I_k}$对待标注图像的支持度指标之一；根据概念格获得近邻图像与待标注图像的语义相关度，计算候选标签集中每个关键词对待标注图像的支持度$\mathrm{sup}(\mathrm{tg_j},\mathrm{I_i})$

$$\mathrm{sup}(\mathrm{tg_j},\mathrm{I_i}) = \mathrm{f}(\mathrm{Z_i}) \times \sum_{\mathrm{k}=1}^{\mathrm{k}} \mathrm{Rel}(\mathrm{I_i},\mathrm{I_k})\varphi(\mathrm{I_k},\mathrm{tg_j}) \quad (3.8)$$

其中，$\varphi(\mathrm{I_k},\mathrm{tg_j})$近邻图像$\mathrm{I_k}$与标签$\mathrm{t_j}$的所属关系，若近邻图像$\mathrm{I_k}$被赋予标签$\mathrm{t_j}$，则$\varphi(\mathrm{I_k},\mathrm{tg_j}) = 1$，反之为 0。得到每个标签词的分数之后，将$\mathrm{sup}(\mathrm{tg_j},\mathrm{I_i})$进行归一化处理，为减少不相关的标签语义词，本文将支持度大于 0.01 的候选标签词保留，去除标签噪声后，作为待标注图像最终的标签词。

参 考 文 献

[1] 胡微微. 基于语义分析的图像多标签标注算法研究[D]. 上海: 华东理工大学, 2013.

[2] 温翔. 弱标注环境下基于多标签深度学习的加速图像标注[D]. 北京:北京交通大学, 2016.

[3] R. Wille. Restructuring Lattice Theory: An Approach Based on Hierarchies of Concepts[J]. Orderd Sets D Reidel,

1982(83): 314-339.

[4] 王亚平, 张素兰, 张继福, 等. 基于模糊概念格的视觉单词生成方法[J]. 小型微型计算机系统, 2016, 37(8):1868-1872.

[5] O. Russakovsky, J. Deng, H. Su, et al. ImageNet Large Scale Visual Recognition Challenge[J]. International Journal of Computer Vision, 2015, 115(3):211-252.