人工智能物联网中面向智能任务的语义通信方法

刘传宏[1]，郭彩丽[1,2]，杨洋[2]，冯春燕[1]，孙启政[1]，陈九九[1]

（1. 北京邮电大学先进信息网络北京实验室，北京 100876；
2. 北京邮电大学网络体系构建与融合北京市重点实验室，北京 100876）

摘 要：随着物联网（IoT）和人工智能（AI）技术的融合发展，传统的数据集中式云计算处理方式难以有效去除数据中大量的冗余信息，给人工智能物联网（AIoT）中智能任务低时延、高精度的需求带来挑战。针对这一挑战，基于深度学习方法提出了 AIoT 中面向智能任务的语义通信方法。针对图像分类任务，在 IoT 设备上利用卷积神经网络（CNN）提取图像的特征图；从语义概念出发，将语义概念和特征图进行关联，提取语义关系；基于语义关系实现语义压缩，减小网络传输的压力，降低智能任务的处理时延。实验和仿真结果表明，对比传统通信方案，所提方法的复杂度仅约为传统方案的 0.8%，同时具有更高的分类任务性能；对比特征图全部传输的方案，所提方法传输时延降低了 80%，大大提升了有效分类准确率。

关键词：物联网；语义通信；图像分类；人工智能；语义压缩

中图分类号：TN929.5

文献标识码：A

DOI: 10.11959/j.issn.1000−436x.2021214

Intelligent task-oriented semantic communication method in artificial intelligence of things

LIU Chuanhong[1], GUO Caili[1,2], YANG Yang[2], FENG Chunyan[1], SUN Qizheng[1], CHEN Jiujiu[1]

1. Beijing Laboratory of Advanced Information Networks, Beijing University of Posts and Telecommunications, Beijing 100876, China
2. Beijing Key Laboratory of Network System Construction and Integration, Beijing University of Posts and Telecommunications, Beijing 100876, China

Abstract: With the integration and development of Internet of things (IoT) and artificial intelligence (AI) technologies, traditional data centralized cloud computing processing methods are difficult to effectively remove a large amount of redundant information in data, which brings challenges to the low-latency and high-precision requirements of intelligent tasks in the artificial intelligence of things (AIoT). In response to this challenge, a semantic communication method oriented to intelligent tasks in AIoT was proposed based on the deep learning method. For image classification tasks, convolutional neural networks (CNN) were used on IoT devices to extract image feature maps. Starting from semantic concepts, semantic concepts and feature maps were associated to extract semantic relationships. Based on the semantic relationships, semantic compression was implemented to reduce the pressure of network transmission and the processing delay of intelligent tasks. Experimental and simulation results show that, compared with traditional communication scheme, the proposed method is only about 0.8% of the traditional scheme, and at the same time it has higher classification task performance. Compared with the scheme that all feature maps are transmitted, the transmission delay of the proposed method is reduced by 80% and the effective accuracy of image classification task is greatly improved.

Keywords: Internet of things, semantic communication, image classification, artificial intelligence, semantic compression

收稿日期：2021−08−03；**修回日期**：2021−11−03

通信作者：郭彩丽，guocaili@bupt.edu.cn

基金项目：国家自然科学基金资助项目（No.92067202）；国家重点研发计划基金资助项目（No.2018YFB1800805）

Foundation Items: The National Natural Science Foundation of China (No.92067202), The National Key Research and Development Program of China (No.2018YFB1800805)

1 引言

随着物联网（IoT, Internet of things）和人工智能（AI, artificial intelligence）技术的融合发展，万物智能互联成为时代所趋，大大激发了人工智能物联网（AIoT, artificial intelligence of things）的繁荣发展[1-4]。在人工智能物联网中，先传输原始数据而后执行智能任务的传统通信方式如图 1 所示，智能设备依靠大量的传感设备（如摄像头等）对外部感知并采集大量数据（如文本、图片等）后，采用传统的信源/信道编码方案，经过调制后将数据集中发送到边/云服务器上；边/云服务器对接收到的信号进行解调和信道/信源解码，得到恢复的原始数据，接着以恢复的数据为输入，利用以深度学习为代表的人工智能技术对数据（如文本、图片等）进行理解和分析，从而完成一系列智能任务[5-7]，如图像分类、目标识别等。在 AIoT 场景中，通信与智能计算深度融合，通信场景从传统的人与人、人与物通信转换为智能体间的通信。通信的目标也不再是准确传输比特数据或者精确传递信号波形，而是准确理解传递的语义信息。这里的语义是指接收者正确理解发送者的信息内容，即对原始数据更精炼的一种“达意”表示[8-9]。因此，传输语义信息的语义通信方法是一种新颖的通信范式，可将原始数据提取出智能计算所需的语义进行传输，有效压缩数据冗余，减少网络传输的压力，降低智能任务的处理时延。

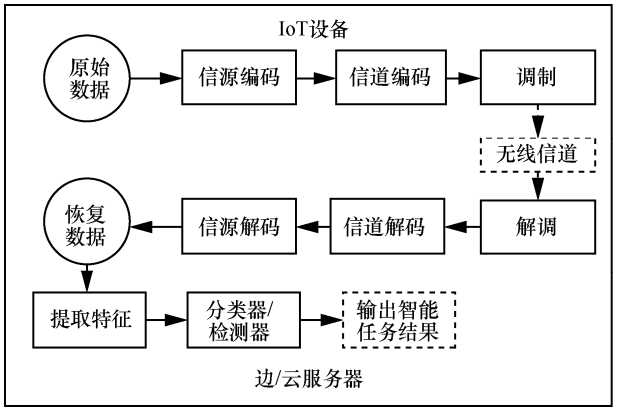


图 1 AIoT 中传统通信方式

目前已经有一些初步的基于深度学习的语义通信方法研究[10-13]。针对文本数据，Xie 等[10]提出了基于 Transformer 的语义通信系统，首次在句子层面明确了语义信息的概念，通过准确恢复句子的语义信息来最大化系统容量和最小化语义误差。在文献[10]的基础上，Xie 等[11]进一步针对物联网场景和无线衰落信道，同时考虑了星座图的设计，提出了轻量化的深度学习语义通信系统，使模型更易于部署在 IoT 设备上。针对图像数据，Bourtsoulatze 等[12]针对无线信道中的图像传输任务，提出了基于卷积神经网络（CNN, convolutional neural network）的联合信源信道编码方法，实现通信系统各编解码模块间的联合优化。Lee 等[13]提出了图像传输和识别联合考虑的通信系统，相较于将通信传输和识别任务分开的传统方法，所提方法可以获得更高的识别精度。总而言之，相较于传统通信方式，语义通信系统可以有效提取并传输数据的语义信息，降低实际传输的数据量，提升通信效率，同时对信道变化具有更强的稳健性[10]。

基于语义通信的方式，将神经网络分开在设备和边/云服务器上联合处理不仅可以充分利用 AIoT 设备端的计算资源，而且可以大大降低待传输的数据量，同时将部分神经网络的参数存储卸载到边/云服务器上，缓解了设备的存储压力。例如，VGG16 网络的参数量约 528 MB，其中全连接层占 75%以上[14]，将全连接层处理放到边/云服务器上，可以降低设备端的存储需求。此外，近些年来人们对用户隐私和数据安全的重视度越来越高，将数据在发送端进行初步特征提取，从而避免了数据的直接传输也可以很好地保护用户隐私和数据安全。

然而，深度神经网络模型复杂，输出的特征数据量较大，在时延敏感的任务中仍然难以满足需求。因此，考虑 AIoT 设备资源和通信资源受限的情况下，如何利用语义关系进一步降低实际传输的数据量和通信系统复杂度，以在缓解通信压力的同时保证智能任务的性能，促进人工智能物联网中通信传输与智能计算 2 个过程的融合成为关键问题。

基于上述考虑，本文基于深度学习方法提出了针对智能任务的语义通信方法，引入神经网络的可解释性，从语义概念（指智能任务中客观表示的某一具体事物，如猫狗分类任务中的猫和狗）出发，将语义概念和特征图（为 CNN 提取的特征向量）进行关联，提取语义关系，在发送端利用提取的语义关系实现语义压缩，大大降低待传输的数据量，使接收端最大程度地理解图片的语义信息，在提升智能任务性能的同时节约通信资源，降低传输时延。本文的主要工作如下。

1) 针对数据传输和智能任务紧密融合的 AIoT 场景，基于深度学习方法提出了一种语义通信系统模型，综合考量了 IoT 设备的能耗和通信资源。

2) 针对图像分类任务，提出了一种语义通信方法。以图像分类任务为导向，关联语义概念和特征图，提取语义关系，并对特征图进行压缩传输，从语义层面实现数据压缩，减少 IoT 设备功耗并降低传输时延。

3) 实验和仿真结果证明了所提人工智能物联网中的语义通信系统方法的可行性，相较于传统通信方法，通信复杂度降低了 99.2%，分类任务性能大大提升；此外，基于所提方法，在不同的特征提取网络下，传输时延降低了约 80%；同时在带宽和时延受限的情况下，提升有效分类准确率可高达约 70%。

2　语义通信系统模型

本节将通信传输和智能任务相结合，基于深度学习方法提出针对智能任务的语义通信系统模型。首先给出人工智能物联网中语义通信流程；然后提出语义通信系统模型，并针对系统模型中的具体模块功能进行详细解释。

图 2 为人工智能物联网中的语义通信流程，实体部分主要由边/云服务器和 IoT 设备组成。基于语义通信的方式，资源受限的 IoT 设备首先利用摄像头采集图像数据，然后利用本地的计算资源完成语义信息提取，接着经语义压缩后将对智能任务重要的语义信息传输到边/云服务器上。边/云服务器平台具有强大的计算能力和内存，将接收到的特征数据输入后续网络进行智能计算，完成智能任务，最终将智能任务结果返回给 IoT 设备。

图 3 为针对智能任务的语义通信系统模型，主要可以分为发送端、物理信道和接收端 3 个部分。发送端功能主要包括语义编码、信道编码和调制，其中语义编码由特征提取、语义关系提取和语义压缩组成；接收端功能主要包括解调、信道解码和智能任务计算（如图像分类任务为分类器，目标检测任务为检测器等）。从图 3 可以看出，所提面向智能任务的语义通信方法将智能任务的处理过程与通信传输过程相融合，先在本地提取语义信息然后传输到接收端直接完成任务，属于先理解后传输的通信方式；而传统通信方法先编码传输原始图片，然后在接收端提取语义信息完成智能任务，属于先传输后理解的通信方式，传输与理解相对独立；此外，语义通信方法利用神经网络实现语义编码，而不再需要传统的信源编码，降低了通信系统的复杂度，这也是传统方法与语义通信方法的另一主要区别。

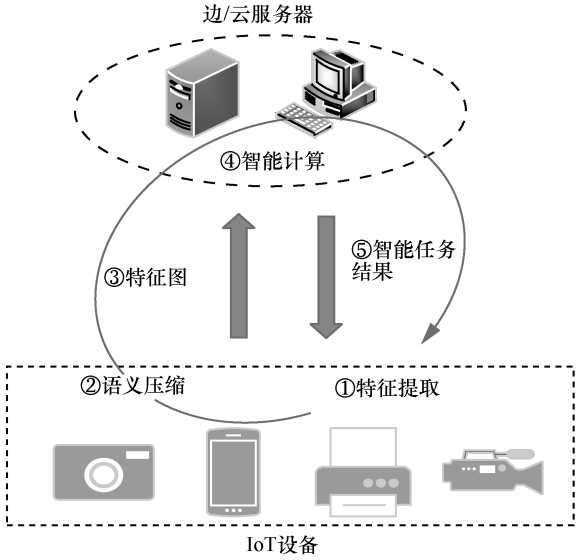


图 2　人工智能物联网中的语义通信流程

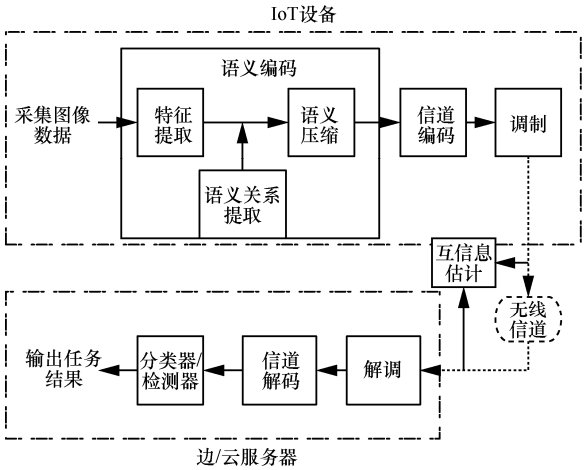


图 3　针对智能任务的语义通信系统模型

IoT 设备利用摄像头采集图像数据，利用 CNN 对图像进行初步特征提取，得到一系列的特征激活图，后续统称为特征图。假设语义通信系统模型的输入是图片 $\boldsymbol{I}$，提取特征的过程可表示为

$$\boldsymbol{A} = S_{\alpha}(\boldsymbol{I}) \tag{1}$$

其中，$S_{\alpha}(\cdot)$ 为特征提取网络，$\boldsymbol{\alpha}$ 为网络参数。

得到经 CNN 提取的特征图后，考虑到对资源紧张的 AIoT 场景的适用性，进一步对实际传输的特征图进行裁剪以实现语义压缩，从而高效完成智能任务。为了实现这一功能，本文提出了语义关系

提取和语义压缩。

语义关系提取以式(1)提取到的特征图为输入，利用每个语义概念对应的神经元激活值对特征图求梯度后取平均，得到特征图对于语义概念的重要性权重。以此为依据，针对具体语义概念对特征图进行排序，得到语义概念和特征图重要性排序之间的关系。针对语义概念 c，对所有特征图进行排序，得到的特征图排序结果为 $\boldsymbol{s}^c$，可表示为

$$\boldsymbol{s}^c = [s_1^c, s_2^c, \cdots, s_N^c], c \in \mathcal{C} \tag{2}$$

其中，$\mathcal{C}$ 为语义概念集合，s_1^c 为对语义概念 c 最重要的特征图索引，s_N^c 为对语义概念 c 最不重要的特征图索引，N 为特征图总数。

语义压缩依据上一步得到的语义概念和特征图排序间的关系，将不重要的部分特征图进行压缩，仅保留重要的语义信息，从而实现去除冗余信息。经语义压缩的输出 $\boldsymbol{X}$ 可表示为

$$\boldsymbol{X} = C_\sigma(\boldsymbol{A}) \tag{3}$$

其中，$C_\sigma(\cdot)$ 为特征图压缩函数，σ 为压缩比例。语义压缩有两大好处：第一，降低了后续的计算资源需求；第二，大大减少了传输的数据量，降低了对通信资源的需求和传输时延。

在本文中，为了主要对比所提通信方法相较于传统通信方法带来的性能增益，暂时不考虑 2 种通信方法中均存在的信道编解码和调制解调部分。因此，可以看作压缩后的特征图直接经无线信道传输，在接收端接收到的特征图数据可表示为

$$\boldsymbol{Y} = h\boldsymbol{X} + n \tag{4}$$

其中，h 为瑞利衰落信道的信道系数，n 为加性白高斯噪声（AWGN, additive white Gaussian noise）。依据香农公式，传输速率表达式如式(5)所示[15]。

$$R = B\mathrm{lb}\left(1 + \frac{h^2 P}{N_0 B}\right) \tag{5}$$

其中，B 为信道带宽，P 为 IoT 设备发送功率，N_0 为 AWGN 功率谱密度。

设计通信系统的主要目标是最大化系统容量，互信息可以用来衡量 2 个变量之间的相关性，为了提高系统容量，降低信道噪声对通信传输过程的影响，提升语义通信系统的稳健性，本文考虑最大化语义通信中信道输入和输出间的互信息。信道输入 $\boldsymbol{X}$ 和输出 $\boldsymbol{Y}$ 之间的互信息计算可表示为

$$I(\boldsymbol{x};\boldsymbol{y}) = \mathrm{E}_{p(\boldsymbol{x},\boldsymbol{y})}\left[\log\frac{p(\boldsymbol{x},\boldsymbol{y})}{p(\boldsymbol{x})p(\boldsymbol{y})}\right] = \mathrm{E}_{p(\boldsymbol{x},\boldsymbol{y})}\left[\log p(\boldsymbol{y}|\boldsymbol{x}) - \log p(\boldsymbol{y})\right] \tag{6}$$

其中，$(\boldsymbol{x},\boldsymbol{y})$ 为输入和输出空间里的随机变量对；$p(\boldsymbol{x})$ 为发送 $\boldsymbol{x}$ 的边缘概率分布，$p(\boldsymbol{y})$ 为接收 $\boldsymbol{y}$ 的边缘概率分布，$p(\boldsymbol{x},\boldsymbol{y})$ 为联合概率分布，$p(\boldsymbol{y}|\boldsymbol{x})$ 为给定 $\boldsymbol{x}$ 的条件下 $\boldsymbol{y}$ 的概率分布。大多情况下，互信息计算困难，但是互信息的上界是可以进行估计的，借鉴文献[16]的思路，利用上界逼近真实互信息，其上界可表示为

$$I_{\mathrm{up}}(\boldsymbol{x};\boldsymbol{y}) := \mathrm{E}_{p(\boldsymbol{x},\boldsymbol{y})}\left[\log p(\boldsymbol{y}|\boldsymbol{x})\right] - \mathrm{E}_{p(\boldsymbol{x})}\mathrm{E}_{p(\boldsymbol{y})}\left[\log p(\boldsymbol{y}|\boldsymbol{x})\right] \tag{7}$$

其证明过程如下。

真实互信息和上界间的差距定义为

$$\begin{aligned}
\Delta &:= I_{\mathrm{up}}(\boldsymbol{x};\boldsymbol{y}) - I(\boldsymbol{x};\boldsymbol{y}) = \\
&\mathrm{E}_{p(\boldsymbol{x},\boldsymbol{y})}\left[\log p(\boldsymbol{y}|\boldsymbol{x})\right] - \mathrm{E}_{p(\boldsymbol{x})}\mathrm{E}_{p(\boldsymbol{y})}\left[\log p(\boldsymbol{y}|\boldsymbol{x})\right] - \\
&\mathrm{E}_{p(\boldsymbol{x},\boldsymbol{y})}\left[\log p(\boldsymbol{y}|\boldsymbol{x}) - \log p(\boldsymbol{y})\right] = \\
&\mathrm{E}_{p(\boldsymbol{x},\boldsymbol{y})}\left[\log p(\boldsymbol{y})\right] - \mathrm{E}_{p(\boldsymbol{x})}\mathrm{E}_{p(\boldsymbol{y})}\left[\log p(\boldsymbol{y}|\boldsymbol{x})\right] = \\
&\mathrm{E}_{p(\boldsymbol{y})}\left[\log p(\boldsymbol{y}) - \mathrm{E}_{p(\boldsymbol{x})}\left[\log p(\boldsymbol{y}|\boldsymbol{x})\right]\right] = \\
&\mathrm{E}_{p(\boldsymbol{y})}\left[\log\left[\mathrm{E}_{p(\boldsymbol{x})}\left[p(\boldsymbol{y}|\boldsymbol{x})\right]\right] - \mathrm{E}_{p(\boldsymbol{x})}\left[\log p(\boldsymbol{y}|\boldsymbol{x})\right]\right] \geqslant 0
\end{aligned} \tag{8}$$

不等号由琴生不等式推导得到。然而变量之间的条件关系是无法直接得到的，因此考虑用带参数 θ 的变分分布 $q_\theta(\boldsymbol{y}|\boldsymbol{x})$ 来近似 $p(\boldsymbol{y}|\boldsymbol{x})$。依据文献[16]的定理 3.2，当满足 $\mathrm{KL}\left(p(\boldsymbol{x},\boldsymbol{y})\middle\|q_\theta(\boldsymbol{x},\boldsymbol{y})\right) \leqslant \mathrm{KL}\left(p(\boldsymbol{x})p(\boldsymbol{y})\middle\|q_\theta(\boldsymbol{x},\boldsymbol{y})\right)$ 条件时，可以得到互信息的上界估计 I_C，因此为了保持上界性质，减少互信息估计之间的差距，目标转化为最小化上界估计，即

$$\begin{aligned}
&\min_\theta \mathrm{KL}\left(p(x)p(y)\middle\|q_\theta(x,y)\right) = \\
&\min_\theta \mathrm{E}_{p(\boldsymbol{x},\boldsymbol{y})}\left[\log\left(p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x})\right) - \log\left(q_\theta(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x})\right)\right] = \\
&\min_\theta \mathrm{E}_{p(\boldsymbol{x},\boldsymbol{y})}\left[\log p(\boldsymbol{y}|\boldsymbol{x})\right] - \mathrm{E}_{p(\boldsymbol{x},\boldsymbol{y})}\left[\log q_\theta(\boldsymbol{y}|\boldsymbol{x})\right]
\end{aligned} \tag{9}$$

其中，$\mathrm{E}_{p(\boldsymbol{x},\boldsymbol{y})}\left[\log p(\boldsymbol{y}|\boldsymbol{x})\right]$ 与 θ 无关，因此可以等价为 $\min_\theta -\mathrm{E}_{p(\boldsymbol{x},\boldsymbol{y})}\left[\log q_\theta(\boldsymbol{y}|\boldsymbol{x})\right]$。

在本文的设计中，变分分布 $q_\theta(\boldsymbol{y}|\boldsymbol{x})$ 通过神经

网络实现，基于大量的变量采样样本对 $(\boldsymbol{x}_i, \boldsymbol{y}_i)$ 最小化对数似然函数。

$$L_{\mathrm{MI}}(\theta) \coloneqq -\frac{1}{N}\sum_{i=1}^{N}\log q_\theta(\boldsymbol{y}_i|\boldsymbol{x}_i) \tag{10}$$

通过梯度下降方法得到 $p(\boldsymbol{y}|\boldsymbol{x})$ 的准确估计。

针对分类任务，接收端将恢复后的特征图 $\boldsymbol{Y}$ 输入分类器网络，输出各类对应的概率值

$$\boldsymbol{p} = Q_{\boldsymbol{\mu}}(\boldsymbol{Y}) \tag{11}$$

其中，$\boldsymbol{p}=[p_1, p_2, \cdots, p_M]$，$p_k$ 为图片分为第 k 类的概率，M 为总类别数，$Q_{\boldsymbol{\mu}}(\cdot)$ 为分类器网络，$\boldsymbol{\mu}$ 为网络参数。

整个分类网络训练过程以最小化交叉熵和最大化互信息为优化目标，损失函数可表示为

$$L(\boldsymbol{y}, \boldsymbol{p}) = -\sum_{i=1}^{M} y_i \log(p_i) - \beta I \tag{12}$$

其中，$\boldsymbol{y}=[y_1, y_2, \cdots, y_M]$ 为样本标签，如果样本的类别是 i，则 $y_i=1$，其余均为 0；β 取值为[0,1]，对互信息项进行加权；I 为信道输入和输出间的互信息。

3　针对图像分类任务的语义通信方法

针对人工智能物联网中的图像分类任务，IoT 设备利用特征提取网络在本地完成图像特征提取。然而，传输全部的特征图仍然对设备功耗和传输时延提出了挑战，针对这一挑战，提出了针对图像分类任务的语义通信方法。本节首先给出针对图像分类任务的语义通信网络结构，然后分别提出图像分类任务导向的语义关系提取方法和语义压缩方法。

3.1　网络结构

图 4 为所提针对图像分类任务的语义通信神经网络结构，以实现高效语义传输并最大程度地提高分类任务的准确度。发送端首先将图像调整为固定尺寸，利用深度 CNN 提取特征，本文采用的 CNN 包括卷积层和池化层，卷积层可以提取输入特征，多层可以逐渐提取更精细的特征。池化层可以降低网络中隐藏层的维数，并降低后续层中的计算量。接着基于提取的语义关系对提取的特征图进行压缩，进一步降低实际传输的数据量。

参考文献[10,17]的工作和结论，一些通信信道可以通过简单的神经网络进行建模，如 AWGN 信道、乘法高斯信道和擦除信道等。在本文中，为了主要关注语义编码，因此考虑 AWGN 信道，并用一层神经网络表示。其物理意义解释如图 5 所示。

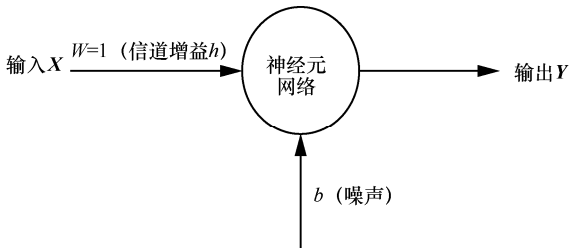


图 5　利用神经网络建模 AWGN 信道

一层神经网络本质上就是输入到输出的一种映射关系，主要由神经元的权重和偏置决定。这里由于已知信道的先验信息即 AWGN 信道，就可以设定神经元的权重 W 为 1，其他信道权重需要相应修改；偏置 b 是添加的高斯随机变量（模拟信道中的高斯噪声），该随机变量的方差对应高斯噪声的功率，主要由信噪比和发送功率决定。即

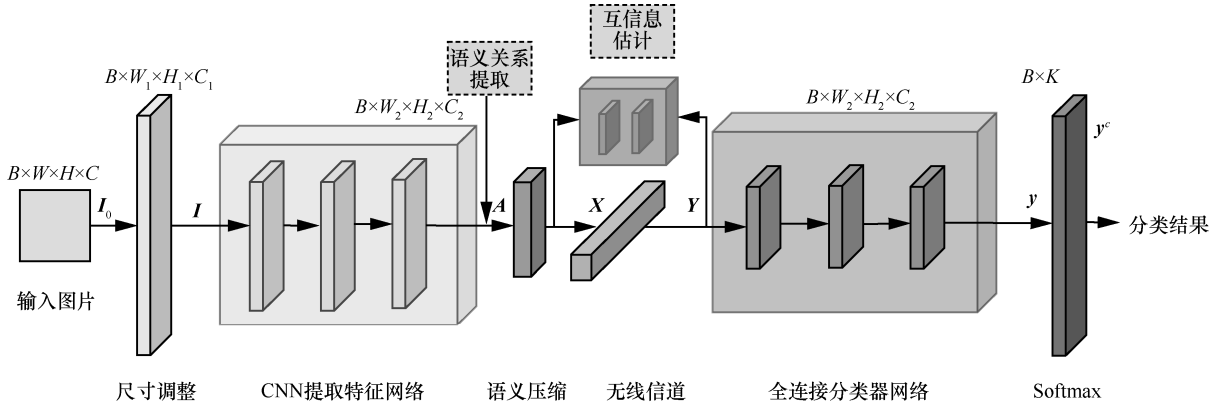


图 4　人工智能物联网中针对图像分类任务的语义通信神经网络结构

$\boldsymbol{Y}=h\boldsymbol{X}+n=\boldsymbol{W}\boldsymbol{X}+b$。

接收端基于三层全连接层构成分类器网络，特征图经分类器及 Softmax 层后输出分类结果。互信息估计网络由两层全连接网络组成。本文和现有语义通信系统的主要区别在于本文将通信传输和图像分类任务相结合，从语义概念出发，利用语义关系进行语义编码和数据压缩，提出了语义关系提取方法和语义压缩方法。

3.2 语义关系提取

本节介绍图像分类任务中基于梯度的语义关系提取方法，将语义概念和特征图进行关联。

如图 4 所示，针对语义概念 c 的得分 $\boldsymbol{y}^c$ 定义为最后一层全连接经过 Softmax 之前的神经元激活值，$\boldsymbol{A}^k$ 为最后一个卷积层的第 k 个特征图，其宽度和高度分别为 w 和 h。首先利用语义概念 c 的得分对第 k 个特征图求梯度，然后经过全局平均池化即可得到针对语义概念 c 的第 k 个特征图的重要性权重 ω_k^c，可表示为

$$\omega_k^c=\frac{1}{wh}\sum_i\sum_j\frac{\partial \boldsymbol{y}^c}{\partial \boldsymbol{A}_{ij}^k} \tag{13}$$

其中，$\boldsymbol{A}_{ij}^k$ 为特征图第 i 行第 j 列处的激活值[18]。针对语义概念 c 的特征图权重向量可表示为

$$\boldsymbol{\omega}^c=[\omega_1^c,\omega_2^c,\cdots,\omega_N^c] \tag{14}$$

相应地，针对所有语义概念得到的特征图权重矩阵可表示为

$$\boldsymbol{\omega}=[\boldsymbol{\omega}^{c_1};\boldsymbol{\omega}^{c_2};\cdots;\boldsymbol{\omega}^{c_K}] \tag{15}$$

其中，$K=|\mathcal{C}|$ 表示语义概念集合的元素个数即语义概念的总数。

针对每一个语义概念，对所有的特征图均进行上述求梯度后取均值的操作，即可得到针对任一语义概念 c 和所有特征图的重要性权重向量 $\boldsymbol{\omega}^c$；紧接着，针对每个语义概念，利用权重值对特征图进行排序，即可得到任一语义概念和特征图重要性排序之间的对应关系，即语义关系[19]，可表示为

$$\boldsymbol{s}=[\boldsymbol{s}^{c_1};\boldsymbol{s}^{c_2};\cdots;\boldsymbol{s}^{c_K}] \tag{16}$$

综上可知，本文提出的基于梯度的语义关系提取方法可以将语义概念和特征图关联起来；针对具体语义概念，将特征图按照重要性进行排序，实现了语义关系的提取，为后续语义压缩提供了依据。

3.3 语义压缩

基于得到的语义关系，可以依据实际的通信环境和资源对特征图进行压缩，选取部分最重要的特征图进行传输，实现语义压缩。

实际应用中，对特征图的压缩准则是至关重要的，显而易见的是，特征图的权重越低即重要性越低，则越可能被压缩且不会对后续结果产生影响。因此，这里的问题转化为如何选择一个合适的压缩阈值。

依据语义概念和特征图重要性排序间的关系，实现特征图压缩，可表示为

$$\boldsymbol{A}^k=\begin{cases}\boldsymbol{A}^k,\omega_k^c\geqslant\omega_n^c\\ 0,\omega_k^c<\omega_n^c\end{cases} \tag{17}$$

其中，ω_n^c 为压缩阈值；n 为针对每个语义概念实际传输的重要特征图个数，可表示为

$$n=\left\lfloor\frac{N(1-\sigma)}{K}\right\rfloor \tag{18}$$

其中，$\sigma\in[0,1)$ 为压缩比例，对应于发送端的语义压缩程度，即 σ 越接近 1，语义压缩程度越大，实际传输特征图的比例为 $1-\sigma$；$\lfloor\bullet\rfloor$ 为向下取整。

压缩的权重阈值可由式(19)得到。

$$\omega_n^c=\boldsymbol{z}(n) \tag{19}$$

其中，$\boldsymbol{z}=\mathrm{sort}([\omega_1^c,\omega_2^c,\cdots,\omega_N^c])$ 为按照权重值从大到小的排序结果，ω_n^c 为第 n 项。

实际传输的特征图中，针对语义概念 c_k 最重要的 n 个特征图索引可表示为

$$\boldsymbol{s}_{\mathrm{T}}^{c_k}=[s_1^{c_k},s_2^{c_k},\cdots,s_n^{c_k}]=\left\{k\middle|\omega_k^{c_k}\geqslant\omega_n^{c_k}\right\} \tag{20}$$

因此，实际经信道传输的特征图索引可表示为

$$\boldsymbol{s}_{\mathrm{T}}=[\boldsymbol{s}_{\mathrm{T}}^{c_1},\boldsymbol{s}_{\mathrm{T}}^{c_2},\cdots,\boldsymbol{s}_{\mathrm{T}}^{c_K}] \tag{21}$$

接着，将压缩后的特征图经无线信道传输。边/云服务器接收特征图数据，输入分类器网络，输出最后的分类结果，并将结果返回给 IoT 设备。

将每张图像的实际传输数据量记为 d，则传输时延可表示为

$$t=\frac{d}{R} \tag{22}$$

设定每张图像数据传输过程中的时延门限为 t_0，当数据实际传输时延大于时延门限（$t>t_0$）时，记作此次传输失败[20]，也就无法完成后续图像分类

任务，判定为分类错误。

人工智能物联网中针对图像分类任务的语义通信算法如算法 1 所示。

算法 1　针对图像分类任务的语义通信算法

输入　待分类图片 $\boldsymbol{I}$，压缩比例 σ=0.902 3

输出　分类结果 $\boldsymbol{p}$

1) 将待分类图片输入特征提取网络，得到特征图 $\boldsymbol{A}$；

2) 基于式(13)计算语义概念激活值对特征图的梯度，得到特征图对语义概念的重要性权重 $\boldsymbol{\omega}$；

3) 针对同一语义概念，对特征图依据重要性进行排序，得到语义概念和特征图排序的对应关系 $\boldsymbol{s}$；

4) 基于式(17)利用语义重要性关系和设定的压缩比例 σ 对特征图进行压缩，得到实际传输的特征图索引向量 $\boldsymbol{s}_{\mathrm{T}}$；

5) 将压缩后的特征图经无线信道传输，计算传输时延 t 并和时延门限 t_0 比较；

6) 若 $t < t_0$，即可成功传输，接收端经分类器输出分类结果并返回；反之，则传输失败，分类错误。

4　实验结果

4.1　实验设置

模型训练和测试数据集为 STL-10 十分类数据集，包含 10 类物体的图片，对应 10 个语义概念，每类 1 300 张图片，500 张训练，800 张测试，每张图片分辨率为 96 dpi×96 dpi。训练过程中，首先将图片扩展到 256×256，然后随机裁剪为 224×224，作为网络的输入。训练和测试环境为 Ubuntu 16.04 + CUDA 10.1，编程语言为 Python，深度学习框架为 Pytorch1.7.0。

为了证明所提方法对于分类网络的普适性，本文选择经典的 CNN 分类网络（VGG16[14] 和 Resnet18[21]）作为特征提取网络，应用所提的语义通信方法进行后续实验仿真。下面分别介绍针对 2 种网络的初始化参数设置。

Resnet18 网络。首先考虑到数据集为十分类任务，调整网络结构最后一层输出神经元个数为 10；接着用在 ImageNet 数据集[22]上预训练好的参数初始化特征提取网络参数，进行后续训练。

VGG16 网络。类似地，首先调整网络结构最后一层输出神经元个数为 10，并将 VGG16 分类器部分两层全连接神经元个数由 4 096 调整为 1 000；接着用在 ImageNet 数据集上训练好的参数初始化特征提取网络参数，训练过程中，冻结前 8 个卷积层参数，以加快训练速度。

本文设计互信息估计网络和分类网络迭代训练，首先，训练互信息估计网络；其次，固定互信息估计网络参数，训练语义通信网络；再次，固定语义通信网络参数，训练互信息估计网络，如此迭代，直到达到收敛条件。

网络训练均使用交叉熵为损失函数，具体训练参数设定如表 1 所示，仿真参数设置如表 2 所示。

表 1　语义通信网络训练参数设置

训练参数	参数取值
Epochs	50
Batchsize	32
优化器	随机梯度下降
学习率	0.01
动量	0.9

表 2　系统仿真参数设置

仿真参数	参数取值
IoT 设备发送功率/W	0.01
噪声功率谱密度/(dBm·Hz^{-1})	−174
射频带宽资源/MHz	11～20
时延门限/ms	0～1.5
待分类图片总数/张	1 000
无线信道	瑞利信道
压缩比例	[0,0.3,0.6,0.8,0.9,0.98]

分类准确率定义为

$$\zeta=\frac{N_s}{N_t} \tag{23}$$

其中，N_t 为待分类的图片总数，N_s 为分类正确的图片数量。

在带宽和时延限制条件下，为了同时考量通信性能和图像分类任务的完成性能，本文提出有效分类准确率为评价指标。将传输和分类分别定义为事件 A 和事件 B，由贝叶斯公式可知，有效分类准确率定义为

$$\begin{aligned} P(B=1)=&P(A=1)P(B=1\big|A=1)+\\ &P(A=0)P(B=1\big|A=0) \end{aligned} \tag{24}$$

其中，$P(A=1)$ 为成功传输的概率，$P(B=1\big|A=1)$ 为成功传输条件下分类成功的概率，$P(A=0)$ 为传输失败的概率，$P(B=1\big|A=0)$ 为传输失败但分类成功

的概率，考虑到传输失败则分类失败，即 $P(B=1|A=0)=0$。则有效分类准确率可以转化为

$$\eta=\beta\gamma \tag{25}$$

其中，β 为给定信噪比和语义压缩程度时的分类准确率 $P(B=1|A=1)$，γ 为给定带宽和时延限制下，实际可以成功传输的图片比例为 $P(A=1)$。

由于本文提出的人工智能物联网中面向智能任务的语义通信方法未改变智能任务的原始神经网络架构，而是在发送端和接收端分别进行部分网络计算，因此并不会增加任务的处理开销，反而由于压缩了部分特征图，可以略微降低接收端的计算量，这也是本文所提方法的优势之一；此外，相较于传统通信方法，所提语义通信方法在发送端利用重要的语义信息替代了原始数据作为存储和传输的主体，利用 CNN 提取语义特征替代了复杂的传统信源编码，因此也并不会增加 IoT 设备的存储和计算压力。后续主要针对分类任务性能进行实验及仿真验证。

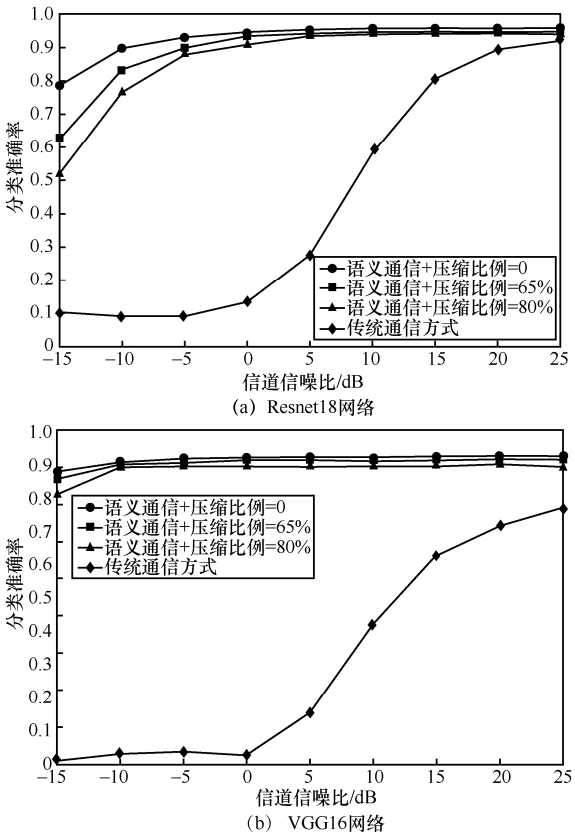
4.2 实验结果分析

本节将对实验和仿真结果进行分析说明，将对比方案命名为传统通信方式，即图像先经过传统 JPEG 编码后传输，接收端经过解码恢复原图，再输入神经网络（Resnet18 或者 VGG16）来完成分类任务。

4.2.1 语义通信和传统通信方式下的分类性能对比分析

为了验证所提语义通信方法相较于传统通信方法更适合应用于 AIoT 场景中，且对信道有更强的稳健性。本节实验对比了所提语义通信方法和传统通信方式的性能，同时验证了在不同信道信噪比条件下，不同语义压缩程度（即不同压缩比例）对最终图像分类任务性能的影响。为了验证所提语义方法对分类网络的普适性，同时基于 VGG16 和 Resnet18 网络进行实验。实验结果如图 6 所示，压缩比例为 0 表示不对特征图进行压缩而全部传输。

图 6(a)和图 6 (b)展示了相同的趋势。基于深度学习的语义通信方式分类任务性能远远好于传统通信方式的分类任务性能，尤其是当信噪比较低时，当信道信噪比为 0 时，传统通信方式几乎无法完成分类，而基于语义通信方式的分类精度可以达到 90%以上。此外，相较于特征图全部传输到接收端，基于语义通信的压缩方式在分类性能上会有一定的损失。然而，从以上 2 个实验结果可以看出，当压缩比例达到 80%时，分类的精度在信噪比大于 0 时损失都很小，均在 2%以内。这说明本文所提的语义通信方法可以实现在几乎不影响任务性能的前提下，大大降低了传输的数据量，降低了传输时延和通信负担，比传统的通信方式更适合于时延和功耗敏感的 IoT 场景中。



(a) Resnet18网络

(b) VGG16网络

图 6 不同网络结构不同压缩比例下不同通信方式下分类准确率随信噪比的变化关系

4.2.2 不同压缩传输方式下的分类任务性能对比分析

为了验证所提语义通信方法在带宽资源紧张的 AIoT 场景中可以有效保障智能任务性能，本节实验对比了不同压缩传输方式下分类性能的差异，对比方案包括：1) 本文所提的语义压缩传输方案；2) 实际场景中的随机压缩传输方案。考虑到实际场景中，带宽资源受限，同时受到时延门限的限制，往往无法将全部特征图经无线信道传输到接收端，因此考虑实际场景中存在的随机丢包，相当于对特征图进行随机压缩后传输。为了更好地比较实验结果，与文献[23]的实验设置保持一

致，这里将带宽定性地定义为满足时延限制条件下允许实际传输的特征图个数，实验结果具体如图 7 所示。

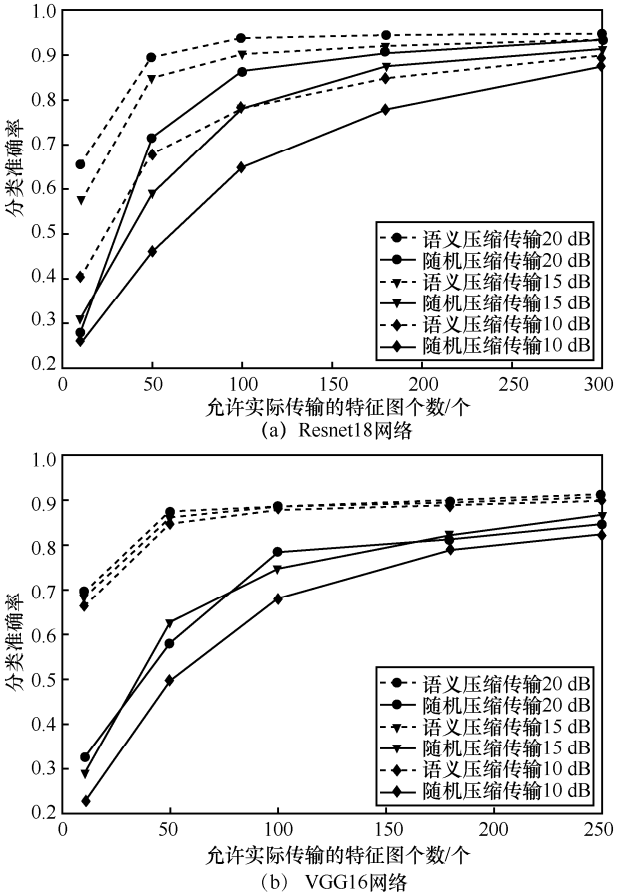


图 7　不同网络结构不同带宽下不同压缩传输方案的分类准确率

图 7(a)和图 7 (b)展示了语义压缩传输方案和随机压缩传输方案下模型分类准确率随着信道带宽的变化关系。2 个实验结果具有相同的趋势，也说明了本文所提方法对于分类网络的普适性。在固定信道带宽和信噪比时，语义压缩传输方案可以提高模型分类准确率，例如在信噪比为 20 dB、允许实际传输的特征图个数为 50 时，语义压缩传输方案相较于随机压缩传输方案可以提高分类准确率约 30%。这是因为本文所提的语义压缩传输在传输过程中，尽可能多地保留了图片重要的语义信息，有利于更好地完成分类任务。同时从实验结果可以看出，随着信道带宽和信噪比的增加，模型的分类准确率也会上升。这是因为无论是传输更多的特征图还是更好的信道条件，都可以保留更多的语义信息，从而有利于分类任务的完成，提高模型分类准确率。

4.2.3　有效分类准确率对比分析

为了验证所提语义通信方法对资源紧张、时延敏感的 AIoT 场景中的适用性，可以有效获得智能任务性能的提升。本节比较了不同带宽和时延限制条件下，基于语义关系的特征图压缩（压缩比例设置为 80%）和传输全部特征图 2 种方式的有效分类准确率，基于 VGG16 分类网络，仿真结果如图 8 所示。

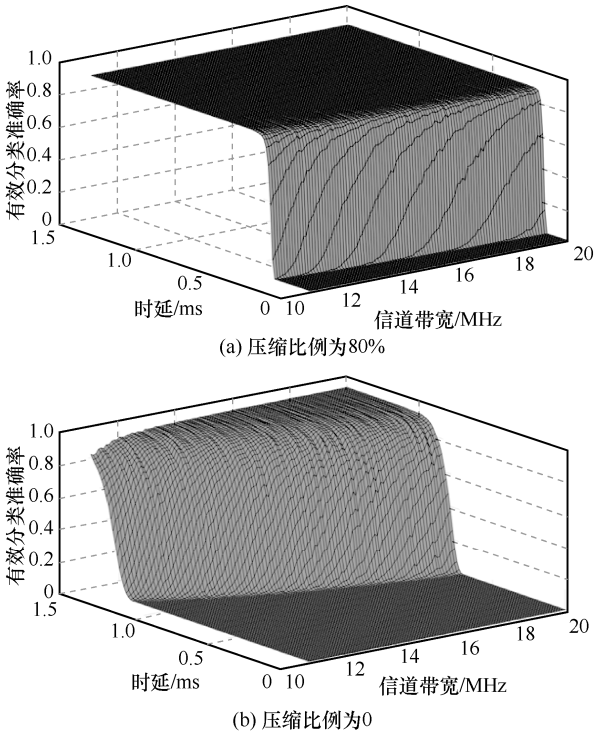


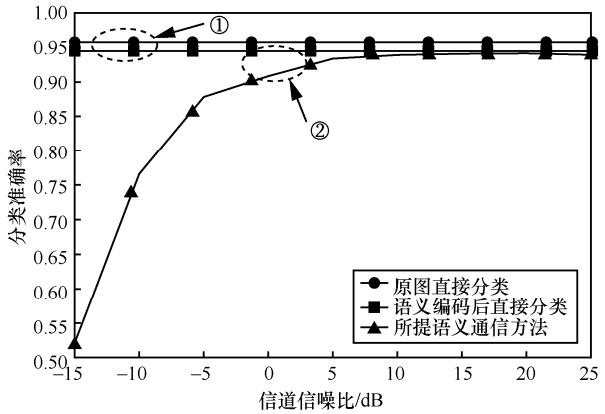
图 8　不同带宽和时延限制下不同压缩比例时的有效分类准确率

从图 8(a)和图 8 (b)对比中可以得出，当信道带宽或者时延门限较小时，特征图完全传输的方式最终几乎无法完成分类任务，这是因为当带宽或时延门限较小时，实际可以完整传输的图片数很少，导致有效分类准确率很低。而经过本文所提方法压缩 80%的特征图后，传输时延也相应降低了 80%，即使在带宽或时延门限较低的条件下，仍然可以有效传输部分重要的卷积核（即图片语义可以得到有效传输），也即成功传输的图片信息比例增大，也提高了有效分类准确率，说明本文所提方法更适合应用在带宽资源紧张和时延要求敏感的 AIoT 场景中，提高模型的有效分类性能。

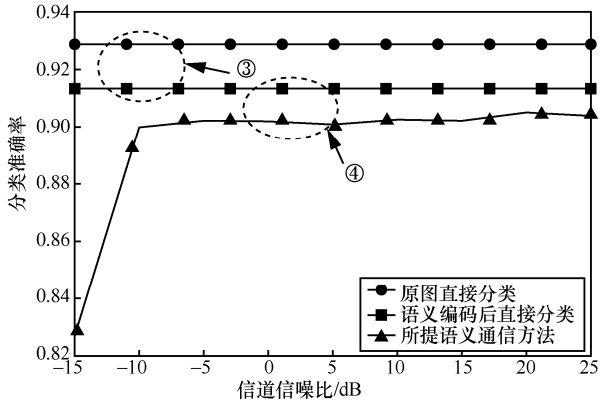
4.2.4　语义压缩及信道传输对任务性能影响分析

为了分别分析所提语义通信方法中语义编码和信道传输 2 部分对智能任务性能的影响，本节基于 2 种分类网络，比较了原图直接分类、语义编码后直接分类和所提语义通信方法（语义编码后经信道传

输再分类）的分类准确率，分析结果如图 9 所示。由于原图直接分类（分类准确率为 0.957 6）和语义编码后直接分类（分类准确率为 0.945 0）均未考虑信道影响，这里的语义编码中的语义压缩比例为 80%，因此图中均为直线，不随信噪比变化。其中原图直接分类和语义编码后直接分类之间的性能差异（图中的①③）是由语义编码引入的；语义编码后直接分类和所提语义通信方法之间的性能差异（图中的②④）是由信道传输（噪声误码等）造成的。从图 9 可以看出，当信道环境较差即信噪比较低时，语义编码直接分类和所提语义通信方法之间的差异较大，这是由于信道传输造成的误差越大，导致性能损失越大；随着信噪比增大，信道中传输造成的性能损失逐渐减小。此外，还可以看出，低信噪比时，信道传输造成的性能损失占主要部分，以基于 Resnet18 网络在−10 dB 时为例，所提语义通信方法的准确率为 0.766 0，则信道传输带来的性能损失为 0.179 0，而语义编码带来的性能损失恒定为 0.012 6；而当信噪比增大到 10 dB 时，信道带来的性能损失降低到了 0.006，甚至小于语义编码带来的性能损失。



(a) Resnet18网络



(b) VGG16网络

图 9 语义编码及信道传输对分类准确率的影响

4.2.5 复杂度对比分析

为了评估所提语义通信方法在 AIoT 场景中的可行性，本节从理论和实验 2 个方面对语义通信方法和传统方法的复杂度对比分析。

1) 理论复杂度对比分析。首先分析所提语义通信网络的计算复杂度。由于所提语义通信网络中计算量主要集中在卷积层，单个卷积层的计算量可表示为 $F\times F\times D_{\text{in}}\times D_{\text{out}}\times W\times H$，其中，$F\times F$ 为滤波器大小，D_{in} 为输入通道数，D_{out} 为滤波器数量，$W\times H$ 为特征图的尺寸，因此语义通信网络的计算复杂度可以近似表示为 $O\left(I_W I_H\right)$，其中 I_W 和 I_H 分别为输入图片的宽和高[12]。传统通信方式计算量主要集中在 JPEG 压缩编解码和分类任务 2 个部分。JPEG 压缩编解码的计算复杂度与像素点数呈线性关系[24]，即 $O(I_W I_H)$。然而传统通信方式的计算复杂度为 JPEG 压缩计算复杂度加上分类任务计算复杂度，远远大于语义通信方法的计算复杂度。

2) 实验复杂度对比分析。本文进一步从实验上比较了面向智能任务的语义通信方法和传统通信方法的复杂度，以单张图片完成分类任务的运行时间来衡量，具体结果如图 10 所示，纵坐标运行时间为对数表示。运行测试程序的计算机的处理器为 Intel(R) Xeon(R) CPU E5-2640 v3 @ 2.60 GHz，16 GB RAM，显卡为 Tesla M40 24 GB。

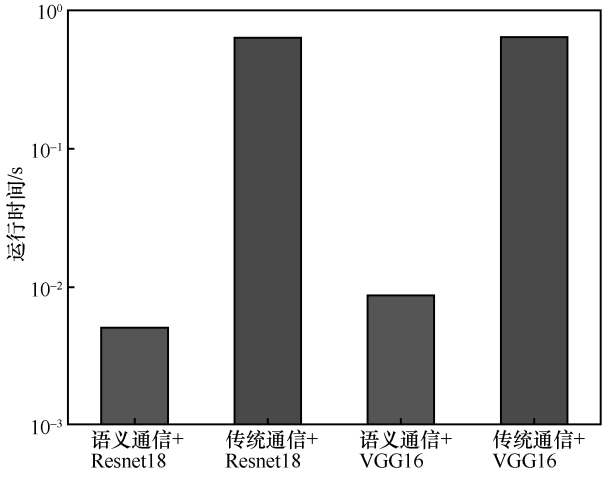


图 10 语义通信和传统方案的复杂度对比

从图 10 可以得出，面向智能任务的语义通信方法所需的运行时间远小于传统通信方法所需的运行时间，而且适用于不同的分类网络，以 Resnet18 网络为例，语义通信方法的运行时间仅为传统方法的 0.8%。这是因为语义通信方法利用分类网络的特

征提取部分实现信源编码，同时利用语义关系实现数据进一步压缩，大大降低了传输和处理时延，因此面向智能任务的语义通信方法相较于传统方法更适合应用于对时延敏感的 AIoT 场景中。本文未将训练时间进行比较是因为针对 AIoT 应用，模型训练通常为离线完成而且仅需要执行一次。

5　结束语

人工智能物联网中，传统的数据集中式云计算处理方式网络传输数据量大，通信时延高，影响智能任务性能。针对这些问题，本文首先基于深度学习方法提出了针对智能任务的语义通信系统模型。特别地，本文研究了人工智能物联网中带宽和时延受限时的图像分类任务，提出了针对图像分类任务的语义通信网络结构，在 IoT 设备上提取图片特征图，基于目标导向的语义关系对特征图进行压缩，有效缓解 IoT 设备的功耗负担，降低通信传输的压力。实验结果验证了本文所提方法的可行性和性能优势。同时顺应通信技术与人工智能紧密结合的大趋势，本文所提方法为众多智能场景下以图像语义理解为主要手段的视觉任务的实现提供了一条新的思路。未来工作将继续研究语义通信方法，考虑不同的信道类型，同时考虑联合语义信道编解码，另外扩展到目标检测等其他视觉任务和其他智能场景中，优化语义提取方法，探究特定环境下最佳的压缩比例，进一步提升模型性能同时降低时延。

参考文献：

[1] AL-FUQAHA A, GUIZANI M, MOHAMMADI M, et al. Internet of things: a survey on enabling technologies, protocols, and applications[J]. IEEE Communications Surveys & Tutorials, 2015, 17(4): 2347-2376.

[2] QIU T, CHEN N, LI K Q, et al. How can heterogeneous Internet of Things build our future: a survey[J]. IEEE Communications Surveys & Tutorials, 2018, 20(3): 2011-2027.

[3] 杨毅宇, 周威, 赵尚儒, 等. 物联网安全研究综述：威胁、检测与防御[J]. 通信学报, 2021, 42(8): 188-205.
YANG Y Y, ZHOU W, ZHAO S R, et al. Survey of IoT security research: threats, detection and defense[J]. Journal on Communications, 2021, 42(8): 188-205.

[4] ÖZYıLMAZ K R, YURDAKUL A. IoT blockchain integration[C]// Security Analytics for the Internet of Everything. Florida: CRC Press, 2020: 29-54.

[5] KE R M, ZHUANG Y F, PU Z Y, et al. A smart, efficient, and reliable parking surveillance system with edge artificial intelligence on IoT devices[J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 22(8): 4962-4974.

[6] 孟倩, 马建峰, 陈克非, 等. 基于云计算平台的物联网加密数据比较方案[J]. 通信学报, 2018, 39(4): 167-175.
MENG Q, MA J F, CHEN K F, et al. Data comparable encryption scheme based on cloud computing in Internet of things[J]. Journal on Communications, 2018, 39(4): 167-175.

[7] GONZÁLEZ G C, NÚÑEZ V E, GARCÍA D, et al. A review of artificial intelligence in the Internet of things[J]. International Journal of Interactive Multimedia and Artificial Intelligence, 2019, 5(4): 9.

[8] RUDOLF C, YEHOSHUA B H. An outline of a theory of semantic information[R]. 1952.

[9] 陈九九, 冯春燕, 郭彩丽, 等. 车联网中视频语义驱动的资源分配算法[J]. 通信学报, 2021, 42(7): 1-11.
CHEN J J, FENG C Y, GUO C L, et al. Video semantics-driven resource allocation algorithm in Internet of vehicles[J]. Journal on Communications, 2021, 42(7): 1-11.

[10] XIE H Q, QIN Z J, LI G Y, et al. Deep learning enabled semantic communication systems[J]. IEEE Transactions on Signal Processing, 2021, 69: 2663-2675.

[11] XIE H Q, QIN Z J. A lite distributed semantic communication system for internet of things[J]. IEEE Journal on Selected Areas in Communications, 2021, 39(1): 142-153.

[12] BOURTSOULATZE E, BURTH K D, GÜNDÜZ D. Deep joint source-channel coding for wireless image transmission[J]. IEEE Transactions on Cognitive Communications and Networking, 2019, 5(3): 567-579.

[13] LEE C H, LIN J W, CHEN P H, et al. Deep learning-constructed joint transmission-recognition for Internet of things[J]. IEEE Access, 2019, 7: 76547-76561.

[14] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv Preprint, arXiv: 1409.1556, 2014.

[15] SHANNON C E. A mathematical theory of communication[J]. Bell system technical journal, 1948, 27(3): 379-423.

[16] CHENG P, HAO W, DAI S, et al. Club: a contrastive log-ratio upper bound of mutual information[C]//International Conference on Machine Learning. Austria: PMLR, 2020: 1779-1788.

[17] FARSAD N, RAO M, GOLDSMITH A. Deep learning for joint source-channel coding of text[C]//Proceedings of 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE Press, 2018: 2326-2330.

[18] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization[J]. International Journal of Computer Vision, 2020, 128(2): 336-359.

[19] FONG R, VEDALDI A. Net2Vec: quantifying and explaining how concepts are encoded by filters in deep neural networks[C]// Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 8730-8738.

[20] QIAO D L, GURSOY M C. Statistical delay tradeoffs in buffer-aided two-hop wireless communication systems[J]. IEEE Transactions on Communications, 2016, 64(11): 4563-4577.

[21] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//Proceedings of 2016 IEEE Conference on

Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2016: 770-778.

[22] DENG J, DONG W, SOCHER R, et al. ImageNet: a large-scale hierarchical image database[C]//Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2009: 248-255.

[23] JANKOWSKI M, GÜNDÜZ D, MIKOLAJCZYK K. Joint device-edge inference over wireless links with pruning[C]//Proceedings of 2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC). Piscataway: IEEE Press, 2020: 1-5.

[24] CHIOU P T, SUN Y, YOUNG G S. A complexity analysis of the JPEG image compression algorithm[C]//Proceedings of 2017 9th Computer Science and Electronic Engineering (CEEC). Piscataway: IEEE Press, 2017: 65-70.

[作者简介]

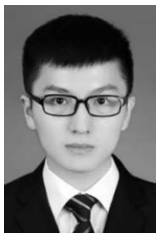
刘传宏（1998− ），男，安徽池州人，北京邮电大学博士生，主要研究方向为深度学习、语义通信、资源分配等。

郭彩丽（1977− ），女，山西太原人，博士，北京邮电大学教授、博士生导师，主要研究方向为语义通信、无线移动通信技术、认知无线电、信号检测与估值、车联网、可见光通信、视觉智能计算、社交跨媒体数据挖掘与分析等。

杨洋（1991− ），男，湖南娄底人，博士，北京邮电大学讲师，主要研究方向为可见光通信、室内定位技术、车联网技术、语义通信技术等。

冯春燕（1963− ），女，北京人，博士，北京邮电大学教授、博士生导师，主要研究方向为无线通信信息传输与处理、宽带通信网络理论与技术、社交网络分析和信息检索、电信大数据分析与挖掘等。

孙启政（1997− ），女，河南安阳人，北京邮电大学博士生，主要研究方向为语义通信、视觉内容理解、深度学习算法等。



陈九九（1994− ），男，湖南平江人，北京邮电大学博士生，主要研究方向为车联网资源分配、语义通信、强化学习算法等。