协作语义通信网络中的图像传输优化

张文静，*IEEE*学生会员，王一宁，*IEEE*学生会员，陈明哲，*IEEE*会员，罗涛，*IEEE*高级会员，以及 Dusit Niyato，*IEEE*会士

arXiv:2301.00433v1 [cs.AI] 1 Jan 2023

摘要

本文构建了一种用于图像数据传输的语义通信框架。在所研究的框架中，一组服务器利用语义通信技术协同向一组用户传输图像数据，使服务器仅传输能够准确捕捉图像含义的语义信息。为评估所研究语义通信系统的性能，本文提出了一种名为图像到图语义相似度（ISS）的多模态度量指标，用于衡量提取的语义信息与原始图像之间的相关性。为满足每个用户的 ISS 要求，每台服务器都必须联合确定待传输的语义信息以及用于语义信息传输的资源块（RB）。由于不同服务器关联用户之间存在同信道干扰，每台服务器都必须与其他服务器协作，以寻找全局最优的语义导向资源块分配。我们将该问题建模为一个优化问题，其目标是在满足 ISS 要求的同时，最小化各服务器平均传输时延之和。为解决该问题，我们提出了一种基于价值分解的熵最大化多智能体强化学习（RL）算法。所提出的算法使每台服务器能够在训练阶段与其他服务器协调，并以分布式方式执行资源块分配，从而以更少的训练迭代次数接近全局最优性能。与传统的多智能体强化学习算法相比，

W. Zhang、Y. Wang 和 T. Luo 就职于北京邮电大学北京先进信息网络实验室，地址为中国北京 100876（电子邮箱：zhangwenjing@bupt.edu.cn；wyy0206@bupt.edu.cn；tluo@bupt. edu.cn）。M. Chen 就职于迈阿密大学电气与计算机工程系及数据科学与计算研究所，地址为美国佛罗里达州科勒尔盖布尔斯 33146（电子邮箱：mingzhe.chen@miami.edu）。D. Niyato 就职于新加坡南洋理工大学计算机科学与工程学院（SCSE）（电子邮箱：dniyato@ntu.edu.sg）。本工作的初步版本 [1] 已被《2022 IEEE国际全球通信会议（GLOBECOM）论文集》接收

所提出的强化学习框架基于对无线环境和语义通信环境的局部观测，提升了服务器对有价值动作的探索能力，以及找到全局最优资源块分配策略的概率。仿真结果表明，与传统的多智能体强化学习算法相比，所提出的算法最多可将传输时延降低 16.1%，并将收敛速度提升 100%。

I. 引言

当前通信技术正努力逼近香农物理容量极限 [2]–[4]。通信与人工智能（AI）技术的融合推动通信向更高层次发展，即从技术层迈向语义层 [5]–[7]。为在无线网络中有效传输快速增长的数据量（即图像、视频和沉浸式数据） [12]–[14]，人们提出了一种名为<风格 id='5'>语义通信</风格>的范式，它将通信技术从以速率为中心转向内容感知 [8]–[11]，。语义通信使设备能够根据原始数据所需表达的含义相互通信，从而提高通信效率 [15]–[17]。然而，当前的语义通信技术主要针对文本和图像数据传输进行研究。与通过词语明确表示语义信息的文本数据相比，图像中的语义信息是隐含的。因此，开发用于图像传输的语义通信框架面临若干挑战，包括：1）以人为中心的语义信息表示；2）图像语义信息的指标设计；3）基于用户服务需求的动态语义信息提取。

A. 相关工作

近年来， [18]–[23]对无线网络中的语义通信进行了研究。在 [18], 中，作者研究了一种基于逻辑概率的语义信息度量方法。在 [19], 中，作者将语义通信系统的语义信道容量定义为语义信息与原始数据之间的互信息。然而， [18] 和 [19] 中设计的两种度量指标仅衡量接收到的语义信息的逻辑真实性，而未考虑语义信息所表达含义的完整性。 [20] 和 [21]中的作者研究了一种基于深度学习的语义通信系统，该系统将原始数据压缩为向量，并将压缩后的向量视为语义

信息。然而，这些向量没有实际含义，人类接收者也无法理解。 [22] 中的作者提出了一种文本语义通信框架，旨在最大化原始数据与语义信息之间的语义相似度。 [23] 中的作者使用接收语义信息的准确性来衡量所提出语义通信系统的性能。然而， [22] 和 [23] 中定义的度量指标基于词级文本数据的一致性，因此无法用于图像数据。

[24]–[27] 中的研究探讨了将语义通信技术用于图像传输。具体而言， [24] 和 [25] 中的研究设计了一种图像语义通信系统，旨在提高图像压缩率。 [26] 中的作者提出了一种图像语义编码模型，并定义了率-感知-失真指标来评估所提模型的性能。 [27] 中的作者研究了一种面向图像的任务驱动语义编码框架。然而， [24]–[27] 中的这些研究将图像的语义信息建模为不可解释的特征向量，无法被人类接收者直接利用和理解。因此， [24]–[27] 中的接收者需要重建原始图像；这种做法效率低下且复杂，因为接收者需要使用神经网络将接收到的数据解读为可解释且有意义的信息。

目前，已有不少研究探讨了将强化学习用于优化语义通信性能。具体而言， [22] 中的作者采用基于注意力的强化学习算法，分析原始数据与其语义信息之间的关系。 [23] 中的作者研究了一种由自批评策略梯度驱动的语义通信系统。 [26]中的研究设计了一种基于强化学习的自适应语义编码模型。 [27] 中的研究利用强化学习确定不同任务中语义编码的量化参数。然而，这些研究未考虑不同智能体之间的协作，因此每个智能体的性能都会受到其他智能体动作的影响，进而降低强化学习所实现的网络性能。 [28] 中的作者采用基于价值分解的深度Q学习网络（DQN），以降低语义通信网络中的传输时延和能耗。然而，由于探索能力较弱，DQN相关的强化学习需要大量用户历史经验才能找到全局最优解。

B. 贡献

本研究的主要目标是设计一种新颖的图像语义通信框架，使一组服务器能够利用语义通信技术协同向用户传输图像。主要贡献包括：

- 我们考虑一种语义通信系统，其中一组服务器协作，使用语义通信技术向一组用户传输图像数据。从图像中提取的语义信息由场景图（SG）建模，该场景图捕捉原始图像中的物体及其关系。

- 为评估语义信息与其对应原始图像之间的语义相似度，我们提出一种全面的多模态图像到图语义相似度（ISS）度量。与结构相似度（SSIM）等衡量像素集合差异的传统度量相比，ISS能够捕捉原始图像与其语义信息之间的含义关联。

- 为满足每位用户的目标ISS要求，每台服务器都必须联合确定要传输的部分语义信息，以及用于语义信息传输的资源块（RB）。我们将这一问题建模为一个优化问题，目标是在满足ISS要求的同时，最小化所有用户的平均传输时延之和。

- 为解决这一优化问题，我们提出一种基于价值分解的熵最大化多智能体深度强化学习（VD-ERL）算法。与传统的多智能体强化学习相比， [28]和 [29], 所提出的算法能够使服务器以较少的训练迭代次数实现全局最优性能。同时，该算法能够增强动作探索能力，提高找到近似最优协作资源块分配策略的概率。

仿真结果表明，与传统的多智能体强化学习算法相比，所提出的 VD-ERL算法可将传输时延最多降低 16.1%，同时将收敛所需的迭代次数减少 50%。据我们所知，这是首项提出图像语义通信框架的研究，该框架联合优化多服务器的资源块分配，在满足 *ISS* 要求的同时，最小化所有用户的平均传输时延之和。

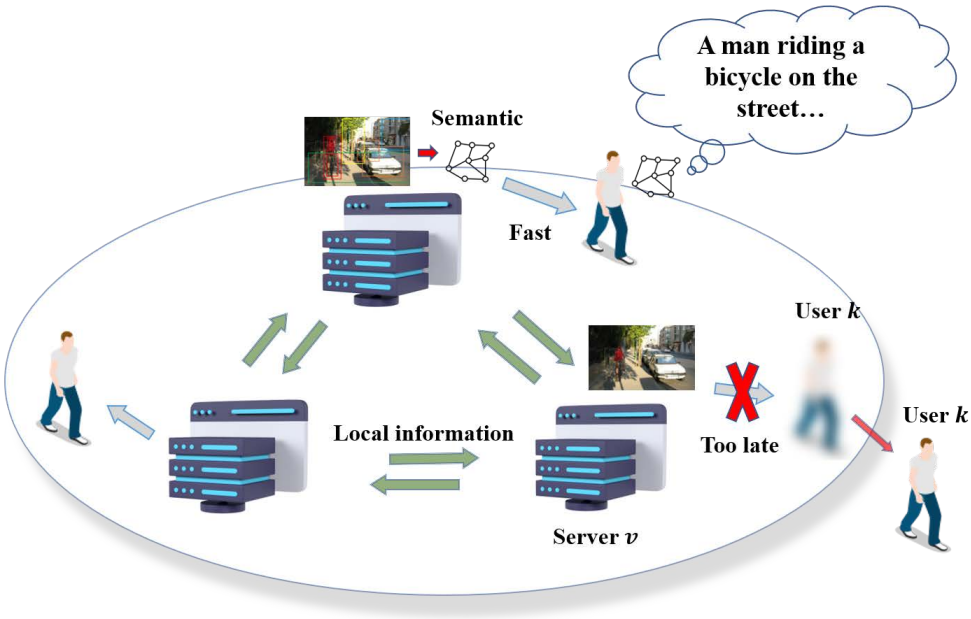


图 1. 协作式多服务器图像语义通信无线网络。

本文其余部分安排如下：第二节介绍所提出的图像语义通信系统模型和问题构建。第三节介绍用于协作式语义导向资源块分配的基于 VD 的熵最大化多智能体强化学习。第四节呈现并讨论数值结果。最后，第五节给出结论。

II. 系统模型与问题构建

考虑一个蜂窝网络，其中一组 $\mathcal{V}$ V 台服务器使用语义通信技术，协同向一组 $\mathcal{U}$ U 个用户传输图像数据，如图1所示。令 $\mathcal{L}_v$表示位于服务器 v服务区域内的用户集合。不同服务器的服务区域可能重叠。如图2所示，每台服务器执行的语义通信过程包含两个阶段：a）语义信息提取；b）语义信息传输。接下来，我们介绍语义信息提取的过程。随后，我们为所提出的图像语义通信框架提出一种多模态度量指标，用于评估原始图像与其提取的语义信息之间的语义相似度。表I总结了本研究中使用的所有参数。

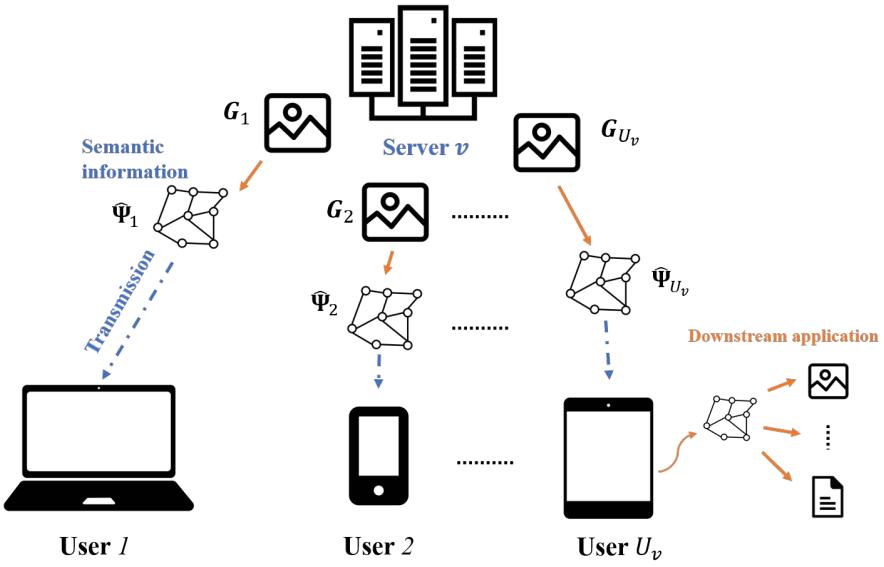


图 2. 每个服务器的图像语义传输框架。

A. 语义信息提取

在我们的模型中，我们假设图像的语义信息由图像中的物体及其关系构成。因此，每幅图像的语义信息都用场景图来建模，该场景图由一组节点和边定义，其中节点表示一个物体（例如，一个男人），边表示两个物体之间的关系，如图3所示。语义三元组是语义信息的基本组成部分，由两个物体及其之间的关系构成。例如，图3中的一个语义三元组为[“男人”], [“骑在……上”],[“自行车”]，其中 [“男人”]和 [“自行车”] 是物体， [“骑在……上”] 是它们之间的关系。服务器需要传输的图像可以用多个语义三元组来描述。

语义信息提取过程分为两个步骤：物体识别和关系捕获。首先，服务器检测物体所在区域并识别其类别。然后，根据物体之间的几何结构和逻辑关联，以及物体的类别，使用深度神经网络模型 [30]–[32]捕获两个物体之间的关系。服务器提取并传输给用户的图像 G_k 语义信息 v k 可以表示为

$$\boldsymbol{\Psi}_{vk} = \left\{ \boldsymbol{\psi}_{vk}^{1}, \boldsymbol{\psi}_{vk}^{2}, \ldots, \boldsymbol{\psi}_{vk}^{n}, \ldots, \boldsymbol{\psi}_{vk}^{N_{vk}} \right\}, \tag{1}$$

表 I：符号列表

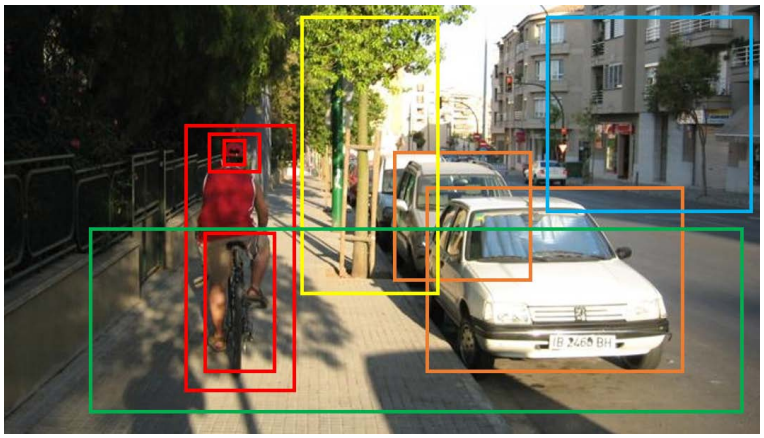
符号	描述	符号	描述
V	服务器数量	U	用户数量
Q	下行正交资源块数量	W	每个资源块的带宽
P	服务器的发射功率	N_0	噪声功率谱密度
I_{vk}^q	资源块 q的干扰	h_{vk}^q	RB q的信道增益
$\boldsymbol{a}_k$	用户 k的RB分配向量	$c_k(\boldsymbol{a}_k)$	用户 k的下行链路容量
G_k	需要传输给用户 k的原始图像	$e_{vk,i}$	G_k中的物体 i
$l_{vk,ij}$	对象间关系$e_{vk,\ i}$ 和 $e_{vk,\ j}$	$\boldsymbol{\Psi}_{vk}$	图像的语义信息 G_k
$\boldsymbol{\psi}_{vk}^n$	$\boldsymbol{\Psi}_{vk}$中的语义三元组 n	$Z(\boldsymbol{\Psi}_{vk})$	$\boldsymbol{\Psi}_{vk}$中的字母数量
$\hat{\boldsymbol{\Psi}}_{vk}$	传输的语义信息	$\hat{N}_{vk}$	其中的语义三元组数量 $\hat{\boldsymbol{\Psi}}_{vk}$
ϵ	语义可靠性阈值	ξ	最低可接受语义相似度
$E\left(\hat{\boldsymbol{\Psi}}_{vk},\ \boldsymbol{a}_{vk}\right)$	图像到图语义相似度	$T\left(\hat{\boldsymbol{\Psi}}_{vk}, \boldsymbol{a}_{vk}\right)$	用户 k的传输时延
$C(G_k)$	向量化图像 G_k	$\boldsymbol{O}_{vk}$	向量化的部分语义信息 $\hat{\boldsymbol{\Psi}}_{vk}$
ρ	关联失败惩罚		

其中，$\boldsymbol{\psi}_{vk}^n = \left(e_{vk,\ i}^n,\ l_{vk,\ ij}^n,\ e_{vk,\ j}^n\right)$ 是一个语义三元组，N_{vk} 是图像 G_k中的语义三元组数量；$e_{vk,\ i}^n$ 是图像 G_k中物体 i 的类别；$l_{vk,\ ij}^n$ 是物体e_{vk}^n之间的关系，$_i$ 和$e_{vk,\ j}^n$。这里，$l_{vk,\ ij}^n$ 具有方向性，因此$l_{vk,\ ij}^n \neq l_{vk,\ ji}^n$。为衡量语义信息的大小，我们将 $Z(\boldsymbol{x})$定义为词序列 $\boldsymbol{x}$ 中的字母数量。因此，每幅图像语义信息 $\boldsymbol{\Psi}_{vk}$ 中的字母总数为

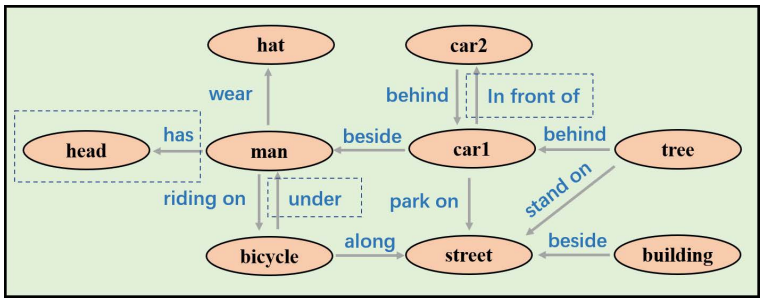
$$Z\left(\boldsymbol{\Psi}_{vk}\right) = \sum_{n=1}^{N_{vk}} \left(Z\left(e_{vk,i}^n\right) + Z\left(l_{vk,ij}^n\right) + Z\left(e_{vk,j}^n\right)\right). \tag{2}$$

例如，在图3中，语义三元组 $\boldsymbol{\psi}_{vk}^n$ = ([“男人”], [“骑在……上”],[“自行车”])中的字母数量为 $Z(\boldsymbol{\psi}_{vk}^n) = Z\left(e_{vk,\ i}^n\right) + Z\left(l_{vk,\ ij}^n\right) + Z\left(e_{vk,\ j}^n\right) = 3 + 8 + 7 = 18$。

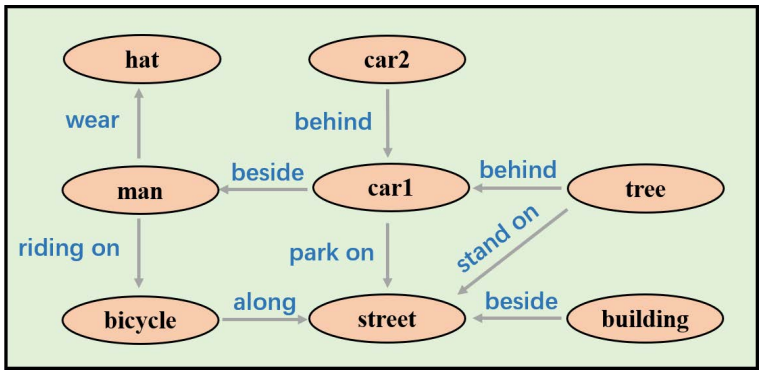
请注意，$\boldsymbol{\Psi}_{vk}$ 中的某些语义三元组可能不包含有用信息。例如，在图3中，我们不希望传输诸如 ([“男人”], [“拥有”],[“头部”]) 这样的无意义语义三元组，也不希望传输诸如 ([“自行车”], [“在……下面”], [“男人”]) 这样的冗余语义三元组。为了



(a) 输入图像。



(b) 提取的语义信息。



(c) 传输的语义信息。

图3. 语义信息提取示例。

提高所考虑的语义通信模型的效率，如图3c）所示，每个服务器 v 都必须传输包含图像最重要含义的语义三元组。服务器 v 传输给用户 k 的部分语义信息可表示为

$$\hat{\boldsymbol{\Psi}}_{vk}=\left\{\hat{\boldsymbol{\psi}}_{vk}^{1},\hat{\boldsymbol{\psi}}_{vk}^{2},\ldots,\hat{\boldsymbol{\psi}}_{vk}^{n},\ldots,\hat{\boldsymbol{\psi}}_{vk}^{\hat{N}_{vk}}\right\}\subset\boldsymbol{\Psi}_{vk},\tag{3}$$

ˆ ˆ其中， N_{vk} 是 $\boldsymbol{\Psi}_{vk}$中选取的语义三元组数量。

B. 传输模型

我们假设采用正交频分多址（正交频分多址）技术。分配一组 $\mathcal{Q}$ 共 Q 个下行正交资源块，为用户提供服务。服务器可以复用所有这些资源块，因此每台服务器可为其关联用户分配 Q 个资源块。服务器向用户 k 传输部分语义信息 $\boldsymbol{\Psi}_{vk}$ 时的下行链路ˆ速率为

$$c_k\left(\boldsymbol{a}_k\right)=\sum_{v=1}^{V}\sum_{q=1}^{Q}a_{vk}^{q}Wlog_2\left(1+\frac{Ph_{vk}^{q}}{I_{vk}^{q}+WN_0}\right),\tag{4}$$

其中， P 是服务器 v的发射功率， W 是资源块 q 的带宽，假设所有资源块的带宽相等；$h_{vk}^{q}=\gamma_{vk}^{q}d_{vk}^{-2}$ 是服务器 v 与用户 k 之间的信道增益，其中 γ_{vk}^{q}为瑞利衰落参数， d_{vk} 为服务器 v 与用户 k之间的距离；$I_{vk}^{q}=\sum_{s\in\mathcal{V}_q,s\neq v}Ph_{s_k}^{q}$ 表示由其他服务器造成的干扰，其中 $\mathcal{V}_q$ 是使用资源块 q的服务器集合； N_0 是噪声功率谱密度； $\boldsymbol{a}_k=[\boldsymbol{a}_{1k},\ldots,\boldsymbol{a}_{vk},\ldots,\boldsymbol{a}_{Vk}]$ 中 $\boldsymbol{a}_{vk}=\left[a_{vk}^{1},\ldots,a_{vk}^{Q}\right]$ 是服务器 v的用户 k的资源块分配向量，其中 $a_{vk}^{q}\in\{0,1\}$为用户-服务器连接索引。具体而言，$a_{vk}^{q}=1$ 表示服务器 v 使用资源块 q向用户 k 传输语义信息；否则为$a_{vk}^{q}=0$。此处，每个用户只能由一台服务器通过一个资源块提供服务，且每台服务器的每个资源块只能分配给一个用户。因此，我们有 $\sum_{v=1}^{V}\sum_{q=1}^{Q}a_{vk}^{q}\leqslant 1,\forall k\in\mathcal{U}$和 $\sum_{k\in\mathcal{U}}a_{vk}^{q}\leqslant 1,\forall v\in\mathcal{V},\forall q\in\mathcal{Q}$。根据式（2）、（3）和（4），服务器 v 向用户 k 传输选取的部分语义ˆ信息 $\boldsymbol{\Psi}_{vk}$ 的传输时延可表示为

$$T\left(\hat{\boldsymbol{\Psi}}_{vk},\boldsymbol{a}_{vk}\right)=\frac{Z\left(\hat{\boldsymbol{\Psi}}_{vk}\right)}{c_{vk}\left(\boldsymbol{a}_{vk}\right)},\tag{5}$$

其中， $c_{vk}(\boldsymbol{a}_{vk})=\sum_{q=1}^{Q}a_{vk}^{q}Wlog_2\left(1+\frac{Ph_{vk}^{q}}{I_{vk}^{q}+WN_0}\right)$是传输速率。这里需要指出的是，我们仅考虑并计算已关联用户的传输时延。由式（5）可知，语义信息的传输时延取决于用户关联、资源块分配以及所传输的部分语义信息的数据大小。因此，对于某个用户，如果其关联的服务器发生变化，那么它从同一幅图像中提取并接收到的语义信息将

有所不同。此外，无线通信环境的变化（如动态信道）也会影响其接收到的语义信息。

C. 图像语义相似度模型

为评估图像语义通信的性能，我们提出了一种称为图像到图语义相似度（ISS）的度量指标。与结构相似度（SSIM）等传统指标不同，后者衡量一组像素之间的差异， [33], 所提出的指标能够捕捉提取的语义信息与其原始图像之间的含义相关性。我们首先使用基于深度神经网络（DNN）的编码器，对原始图像 G_k以及用户 k接收到的语义信息$\hat{}\,\boldsymbol{\Psi}_{vk}$ 进行向量化。向量化后的图像数据为 $\boldsymbol{C}(G_k)$，向量化后的语义信息为 $\boldsymbol{O}_{vk} = \left\{\boldsymbol{C}\left(\hat{}\,\boldsymbol{\psi}_{vk}^{1}\right), \ldots, \boldsymbol{C}\left(\hat{}\,\boldsymbol{\psi}_{vk}^{n}\right), \ldots, \boldsymbol{C}\left(\hat{}\,\boldsymbol{\psi}_{vk}^{N_{vk}}\right)\right\}$，其中$\boldsymbol{C}(\cdot)$是向量化函数，它通过匹配含义相似的文本-图像对，构建输入语义信息与图像之间的关系。

所提出的 ISS 指标定义为图像向量与其归一化语义三元组向量之间的夹角余弦，该值通过将图像向量投影到语义信息向量集上计算得到。为构建语义信息向量集的基，使用格拉姆-施密特算法对语义信息向量进行正交化，具体如下给出 $\boldsymbol{O}_{vk} = \{\boldsymbol{C}(\hat{}\,\boldsymbol{\psi}_{vk}^{1}) \ldots, \boldsymbol{C}(\hat{}\,\boldsymbol{\psi}_{vk}^{n}) \ldots, \boldsymbol{C}(\hat{}\,\boldsymbol{\psi}_{vk}^{N_{vk}})\}$。随后，将从服务器 v 传输到用户 k 的语义信息$\hat{}\,\boldsymbol{\Psi}_{vk}$ 的 ISS 定义为:

$$E\left(\hat{\boldsymbol{\Psi}}_{vk}, \boldsymbol{a}_{vk}\right) = \left(\sum_{q=1}^{Q} a_{vk}^{q}\right) \frac{\| \sum_{n=1}^{\hat{N}_{vk}} |\overline{\boldsymbol{C}\left(\hat{\boldsymbol{\psi}}_{vk}^{n}\right)} \cdot \boldsymbol{C}\left(G_k\right)^{T}| \overline{\boldsymbol{C}\left(\hat{\boldsymbol{\psi}}_{vk}^{n}\right)} \|}{\|\boldsymbol{C}\left(G_k\right)\|}. \tag{6}$$

由式（6）可知，随着传输的语义三元组数量增加，ISS 的值也随之增加，这与人类认知的客观规律一致 [34]。

在所提出的框架中，每个服务器 v 仅传输部分语义信息，因此，接收到的语义信息仅包含图像的一部分含义。我们将每个用户的最低可接受图像到图语义相似度定义为 ξ。随后，将接收到的部分语义信息满足 $E\left(\hat{}\,\boldsymbol{\Psi}_{vk}, \boldsymbol{a}_{vk}\right) \geqslant \xi$ 的概率定义为语义可靠性，其表示为

$$P\left(E\left(\hat{\boldsymbol{\Psi}}_{vk}, \boldsymbol{a}_{vk}\right) \geqslant \xi\right) \geqslant \epsilon, \tag{7}$$

其中， ϵ 是语义可靠性阈值，用于调整可靠语义传输的概率。例如， $\xi = 0.6$ 和 $\epsilon = 0.9$ 表示至少 90% 的语义信息传输必须满足 $E\left(\hat{\boldsymbol{\Psi}}_{vk}, \boldsymbol{a}_{vk}\right) \geqslant 0.6$。

D. 问题构建

给定已定义的系统模型，我们的目标是在满足语义可靠性要求的同时，最小化所有用户的平均传输时延。该最小化问题包括优化用户关联、资源块分配，以及确定要传输的部分语义信息。平均传输时延最小化问题表述如下：

$$\min_{\hat{\boldsymbol{\Psi}}_{vk}, \boldsymbol{a}_{vk}} \frac{\sum_{v\in\mathcal{V}} \sum_{k\in\mathcal{U}_v} T\left(\hat{\boldsymbol{\Psi}}_{vk}, \boldsymbol{a}_{vk}\right)}{\sum_{v\in\mathcal{V}} |\mathcal{U}_v|} \tag{8}$$

$$\text{s.t.} \quad a_{vk}^q \in \{0,1\}, \forall k \in \mathcal{U}_v, \forall v \in \mathcal{V}, \forall q \in \mathcal{Q}, \tag{8a}$$

$$\sum_{v\in\mathcal{V}} \sum_{q\in\mathcal{Q}} a_{vk}^q \leqslant 1, \forall k \in \mathcal{U}_v, \tag{8b}$$

$$\sum_{k\in\mathcal{U}_v} a_{vk}^q \leqslant 1, \forall v \in \mathcal{V}, \forall q \in \mathcal{Q}, \tag{8c}$$

$$\mathcal{U}_v \subset \mathcal{L}_v, \forall v \in \mathcal{V}, \tag{8d}$$

$$P\left(E\left(\hat{\boldsymbol{\Psi}}_{vk}, \boldsymbol{a}_{vk}\right) \geqslant \xi\right) \geqslant \epsilon, \hat{\boldsymbol{\Psi}}_{vk} \subset \boldsymbol{\Psi}_{vk}, \forall k \in \mathcal{U}_v, \tag{8e}$$

其中， $\mathcal{U}_v$ 是与服务器 v 关联的用户集合， $\mathcal{L}_v$ 是位于服务器 v服务区域内的用户集合。约束(8a)、(8b)和(8c)确保每台服务器都可以为每个关联用户分配一个资源块，且在图像语义信息传输过程中，每个资源块只能由一个用户占用。约束(8e)表示每个用户的语义可靠性要求。由于约束(8e)是非凸的，且语义信息提取依赖于深度神经网络模型，因此无法在多项式时间内通过传统优化算法求解问题(8)。此外，单台服务器无法观测全局无线通信环境，也无法获取与其他服务器关联的用户信息。因此，集中式强化学习算法（例如 DQN）只能基于局部观测，最小化所部署服务器的传输时延。为求解旨在最小化所有用户平均传输时延之和的问题(8)，我们引入一种多智能体

强化学习算法，使所有服务器能够协同优化所研究语义通信网络的资源分配。

III. 基于价值分解的熵最大化多智能体强化学习方法

为有效解决问题（8），我们引入一种基于价值分解的 [35] 熵最大化多智能体强化学习（VD-ERL）算法，以最小化所有服务器的平均传输时延，而不是单个服务器的平均传输时延。我们首先介绍所提出的 VD-ERL 方法的组成部分，然后介绍所提出的 VD-ERL 方法的训练过程。

A. VD-ERL 方法的组成部分

本节介绍所提出的 VD-ERL 方法的基本组成部分如下：

- 智能体：智能体是负责确定资源块分配以及需要向其关联用户传输的语义三元组集合的服务器。
- 状态：状态定义为 $\boldsymbol{s}=[\boldsymbol{s}_1,\ldots,\boldsymbol{s}_v,\ldots,\boldsymbol{s}_V]$，其中 $\boldsymbol{s}_v=[\boldsymbol{\gamma}_v,\boldsymbol{\beta}_v]$ 表示服务器 v的部分状态。 $\boldsymbol{\gamma}_v=\left[\gamma_v^1,\ldots,\gamma_v^Q\right]$ 是可用资源块向量，其中 $\gamma_v^q=0$ 表示资源块 q 已被分配，$\gamma_v^q=1$表示未被分配。 $\boldsymbol{\beta}_v=\left[\boldsymbol{\beta}_{v_1},\ldots,\boldsymbol{\beta}_{vk},\ldots,\boldsymbol{\beta}_{v_{|\mathcal{L}_v|}}\right]$是位于服务器 v 覆盖范围内的用户的语义三元组评分矩阵，用于评估语义可靠性，其中 $|\mathcal{L}_v|$ 是服务器 v服务区域内的用户数量， $\boldsymbol{\beta}_{vk}=\left[\beta(\boldsymbol{\psi}_{vk}^1)\ \ldots,\ \beta(\boldsymbol{\psi}_{vk}^n)\ \ldots,\ \beta\left(\boldsymbol{\psi}_{vk}^{N_{vk}}\right)\right]$ 是语义信息 $\boldsymbol{\Psi}_{vk}$中所有语义三元组的评分向量。每个语义三元组 $\boldsymbol{\psi}_{vk}^n$的评分可表示为

$$\beta\left(\boldsymbol{\psi}_{vk}^n\right)=\frac{\exp\left(\mu\left(e_{vk,i}^n\right)\mu\left(l_{vk,ij}^n\right)\mu\left(e_{vk,j}^n\right)\right)}{\sum_{n=1}^{N_{vk}}\exp\left(\mu\left(e_{vk,i}^n\right)\mu\left(l_{vk,ij}^n\right)\mu\left(e_{vk,j}^n\right)\right)},\tag{9}$$

其中，$\mu\left(e_{vk,i}^n\right)$ 是从图像 G_{vk} 中检测到物体 $e_{vk,i}^n$ 的概率，而 $\mu\left(l_{vk,ij}^n\right)$是在给定物体 $e_{vk,i}^n$ 和 $e_{vk,j}^n$的情况下推断出关系 $l_{vk,ij}^n$ 的条件概率。在式 (9) 中，$\mu\left(e_{vk,i}^n\right)$ 和 $\mu\left(l_{vk,ij}^n\right)$ 可以通过场景图生成模型 [32]获得。每个语义三元组 $\boldsymbol{\psi}_{vk}^n$ 的得分表示从原始图像中提取三元组 $\boldsymbol{\psi}_{vk}^n$ 的概率，并将用于选择要传输的部分语义信息。具体而言，得分较高的语义三元组对语义信息的贡献更大

在此，我们指出，每个服务器 v 只能观察到其部分状态 $\boldsymbol{s}_v$。

- 动作：每个服务器的动作 $\boldsymbol{\alpha}_v$ v 均为资源块分配，其定义如下：

$$\boldsymbol{\alpha}_v = \left[\boldsymbol{a}_{v1}, \ldots, \boldsymbol{a}_{vk}, \ldots, \boldsymbol{a}_{v|\mathcal{L}_v|}\right], \tag{10}$$

其中，表示 RB 分配向量的 $\boldsymbol{a}_{vk}$ 是问题 (8) 中的变量。所有分布式服务器的动作向量为 $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_v, \ldots, \boldsymbol{\alpha}_V]$。

- 策略：策略是每个智能体在给定部分状态 $\boldsymbol{s}_v$ 下选择动作 $\boldsymbol{\alpha}_v$ 的条件概率。策略由参数为 $\boldsymbol{\phi}_v$ 的深度神经网络实现，该网络建立语义三元组分数、ISS 与每个用户的传输时延之间的关系。于是，每个智能体在给定部分状态 $\boldsymbol{s}_v$ 下采取动作 $\boldsymbol{\alpha}_v$ 的条件概率可表示为 $\boldsymbol{\pi}_{\boldsymbol{\phi}_v}(\boldsymbol{\alpha}_v \mid \boldsymbol{s}_v)$。为提高动作探索能力，策略网络的训练目标不仅是最大化期望奖励，还要最大化动作 $\mathcal{H}(\boldsymbol{\pi}_{\boldsymbol{\phi}_v}(\boldsymbol{\alpha}_v \mid \boldsymbol{s}_v))$ 的熵，从而促使智能体以更随机的方式选择动作。

- 奖励：每个服务器的奖励用于衡量所选动作在语义可靠性和传输时延方面带来的收益。为了计算每个服务器的奖励$\hat{}v$，我们首先需要确定将传输给用户 k 的部分语义信息 $\boldsymbol{\Psi}_{vk}$。具体而言，根据状态 $\boldsymbol{s}_v$ 和动作 $\boldsymbol{\alpha}_v$，我们可以按照分数向量 $\boldsymbol{\beta}_{vk}$对语义三元组 $\boldsymbol{\Psi}_{vk}$ 进行排序。具体而言，排序后的语义三元组向量为 $\left[\hat{}\boldsymbol{\psi}_{vk}^1, \ldots\ldots, \hat{}\boldsymbol{\psi}_{vk}^{N_{vk}}\right]$，其中$\hat{}\boldsymbol{\psi}_{vk}^1$是分数最高的三元组，而$\hat{}\boldsymbol{\psi}_{vk}^{N_{vk}}$是分数最低的三元组。给定这个排序后的三元组向量，我们使用迭代算法选择若干三元组，以满足约束条件 (8e)，同时尽量缩短传输时间。用于确定所选三元组 $\left[\hat{}\boldsymbol{\psi}_{vk}^1, \ldots\ldots, \hat{}\boldsymbol{\psi}_{vk}^n\right]$ 以生成语义信息的迭代算法总结见**算法 1**。

然后，与服务器 v 关联的每个用户 k 的奖励为

$$r_{vk}\left(\boldsymbol{s}_v, \boldsymbol{a}_{vk}\right) = \sum_{q=1}^{Q} a_{vk}^q \left[\mathbb{1}_{\left\{E\left(\hat{\boldsymbol{\Psi}}_{vk}, \boldsymbol{a}_{vk}\right) \geqslant \xi\right\}} \cdot \eta - T\left(\hat{\boldsymbol{\Psi}}_{vk}, \boldsymbol{a}_{vk}\right)\right] \tag{11}$$

其中，η是常数偏置，$T\left(\hat{}\hat{\boldsymbol{\Psi}}_{vk}, \boldsymbol{a}_{vk}\right)$是传输时延，而$\mathbb{1}_{\{E(\hat{\boldsymbol{\Psi}}_{vk}, \boldsymbol{a}_{vk}) \geqslant \xi\}}\hat{}$是一个函数，用于指示接收到的语义信息 $\boldsymbol{\Psi}_{vk}$ 是否满足约束（8e）中定义的语义可靠性要求。

算法1 语义三元组选择算法。

1：**输入：**语义三元组分数 $\boldsymbol{\beta}_{vk}$、语义三元组数量 N_{vk}以及最小语义相似度 ξ。2：**初始化：**按照 $\boldsymbol{\beta}_{vk}$对 $\boldsymbol{\Psi}_{vk}$ 中的语义三元组按降序排序。3：**对于** $n=1 \rightarrow N_{vk}$ **执行**4：选择得分最高的 n 个三元组$\hat{}\boldsymbol{\Psi}_{vk}=\left[\hat{}\boldsymbol{\psi}_{vk}^{1}, \ldots, \hat{}\boldsymbol{\psi}_{vk}^{n}\right]$。5：根据语义三元组分数 $\widetilde{E}\left(\hat{}\boldsymbol{\Psi}_{vk}\right)=\sum_{i=1}^{n} \boldsymbol{\beta}\left(\hat{}\boldsymbol{\psi}_{vk}^{i}\right)$估算语义相似度。6：**如果** $\widetilde{E}\left(\hat{}\boldsymbol{\Psi}_{vk}\right) \geqslant \xi$**则**7：**结束循环**8：**结束条件判断**9：**结束循环**$\hat{}$10：**输出：**选定的语义三元组 $\boldsymbol{\Psi}_{vk}$。

由于服务器以分布式方式将RB资源分配给用户，且每台服务器都不知道其他服务器的RB分配方案，因此多台服务器可能会将资源块分配给同一用户，而该用户只能使用其中一台服务器的资源块，从而浪费其他服务器的资源块。为了提高资源块利用率，我们在奖励函数中加入负惩罚 ρ 。具体而言，所有服务器的总奖励表示为

$$\begin{aligned} r\left(\boldsymbol{s}, \boldsymbol{\alpha}\right) &= \sum_{v=1}^{V} r_{v}\left(\boldsymbol{s}_{v}, \boldsymbol{\alpha}_{v}\right) \\ &= \sum_{v=1}^{V} \sum_{k \in \mathcal{U}_{v}}\left[\mathbb{1}_{\left\{\sum_{\zeta \neq v, \zeta \in \mathcal{V}} \sum_{q=1}^{Q} \boldsymbol{a}_{vk}^{q}=0\right\}} r_{vk}\left(\boldsymbol{s}_{v}, \boldsymbol{a}_{vk}\right)+\left(1-\mathbb{1}_{\left\{\sum_{\zeta \neq v, \zeta \in \mathcal{V}} \sum_{q=1}^{Q} \boldsymbol{a}_{vk}^{q}=0\right\}}\right) \rho\right], \end{aligned} \tag{12}$$

其中，$r_{v}\left(\boldsymbol{s}_{v}, \boldsymbol{\alpha}_{v}\right)$是服务器 v 的奖励，$\mathbb{1}_{\left\{\sum_{\zeta \neq v, \zeta \in \mathcal{V}} \sum_{q=1}^{Q} \boldsymbol{a}_{vk}^{q}=0\right\}}$ 是一个用于指示用户 k 是否由其他服务器服务的函数。根据式（12）可知，当某个资源块利用不足时，奖励将为 ρ。

- 个体Q值函数：每个服务器 v 的个体Q值函数定义为 $Q_{\boldsymbol{\theta}_{v}}\left(\boldsymbol{s}_{v}, \boldsymbol{\alpha}_{v}\right)$，用于估计在给定服务器 v 的部分状态$\boldsymbol{s}_{v}$ 和所选动作 $\boldsymbol{\alpha}_{v}$ 下的期望奖励。每个服务器 v 使用参数为 $\boldsymbol{\theta}_{v}$ 的深度神经网络来近似个体 Q 值函数。由于每个服务器只能观测其服务区域内用户的状态，因此每个服务器都会将其个体Q值发送给其他服务器，以便估计全局Q值函数，相关内容将在

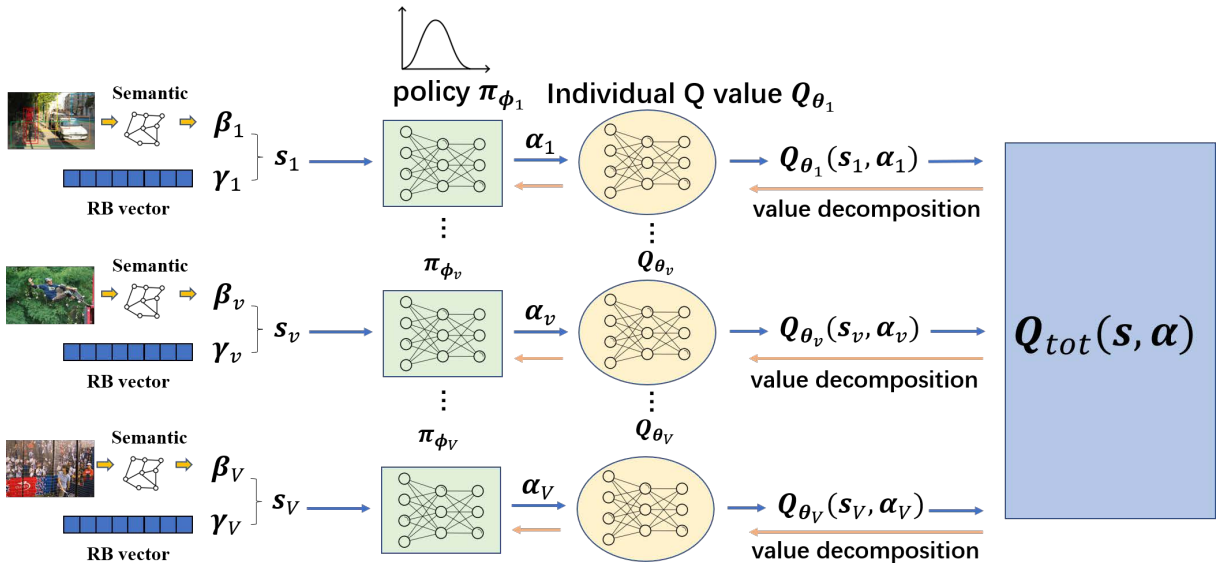


图4. 所提出的 VD-ERL算法的训练过程。

下一条要点中说明。

- 全局Q值函数：全局Q值函数定义为 $Q_{tot}(\boldsymbol{s}, \boldsymbol{\alpha})$，用于估计所有分布式服务器的总期望奖励。对于所提出的 VD-ERL 算法，我们假设所有服务器的全局 Q 值等于各服务器个体Q值之和，即 [35]

$$Q_{tot}(\boldsymbol{s}, \boldsymbol{\alpha}) = \sum_{v=1}^{V} Q_{\boldsymbol{\theta}_v}(\boldsymbol{s}_v, \boldsymbol{\alpha}_v). \tag{13}$$

每个服务器 v 的目标是通过训练其策略网络，协同最大化总期望奖励，即最大化全局 Q 值。训练完成后，每个服务器都可以基于全局Q值函数找到最优策略，从而在满足所有用户语义可靠性要求的同时，最小化所有用户的传输时延总和。

B. 面向语义的资源分配的 *VD-ERL*算法

接下来，我们介绍服务器如何使用所提出的 VD-ERL算法，协同最小化平均语义信息传输时延之和。首先，每个智能体都会收集包含部分状态 $\boldsymbol{s}_v$ 和动作 $\boldsymbol{\alpha}_v$的局部信息。然后，每个智能体将其局部信息传输给其他智能体，以计算其服务器奖励 $r_v(\boldsymbol{s}_v, \boldsymbol{\alpha}_v)$以及

所有服务器的总奖励 $r(\boldsymbol{s},\boldsymbol{\alpha}) = \sum_{v=1}^{V} r_v(\boldsymbol{s}_v,\boldsymbol{\alpha}_v)$。最后，如图4所示，每个智能体根据总奖励和全局 Q 值函数更新其策略网络和个体Q值函数。具体而言，每个智能体首先根据当前策略 $\boldsymbol{\pi}_{\boldsymbol{\phi}_v}\left(\boldsymbol{s}_v^d,\boldsymbol{\alpha}_v^d\right)$收集一组轨迹 $\mathcal{D}_v = \left\{\boldsymbol{\tau}_v^1,\ldots,\boldsymbol{\tau}_v^d,\ldots,\boldsymbol{\tau}_v^D\right\}$。然后，每个智能体从 $\mathcal{D}_v$ 中抽取一定批量大小的轨迹，并计算总奖励和全局 Q 值，以训练个体Q值函数 $Q_{\boldsymbol{\theta}_v}$ 和策略网络 $\boldsymbol{\pi}_{\boldsymbol{\phi}_v}$。最后，每个服务器在给定状态 $\boldsymbol{s}_v$ 下，根据更新后的策略网络 $\boldsymbol{\pi}_{\boldsymbol{\phi}_v}$ 采样动作 $\boldsymbol{\alpha}_v$，以收集用于下一次迭代的新轨迹。全局Q值函数的损失函数 $Q_{tot}(\boldsymbol{s},\boldsymbol{\alpha})$定义如下

$$J(\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_V) = \mathbb{E}_{\boldsymbol{\tau}^d\in\mathcal{D}}\left[Q_{tot}(\boldsymbol{s},\boldsymbol{\alpha}) - r(\boldsymbol{s},\boldsymbol{\alpha}) - \max_{\boldsymbol{\alpha}'} Q_{tot}(\boldsymbol{s}',\boldsymbol{\alpha}')\right]^2, \tag{14}$$

其中，$\max_{\boldsymbol{\alpha}'} Q_{tot}(\boldsymbol{s}',\ \boldsymbol{\alpha}')$ 是下一状态 $\boldsymbol{s}'$ 的最大全局 Q 值。随着每个个体Q值的增加，全局 Q 值也会单调增加；也就是说，具有较高个体Q值的智能体所采取的动作对整个无线网络也有价值。因此，每个服务器训练其个体Q值函数的目标是最大化全局 Q 值。每个服务器的个体Q值函数 $Q_{\boldsymbol{\theta}_v}(\boldsymbol{s}_v,\ \boldsymbol{\alpha}_v)$ 可通过以下梯度下降法进行更新：

$$\boldsymbol{\theta}_v \leftarrow \boldsymbol{\theta}_v - \lambda_{\boldsymbol{\theta}_v}\nabla_{\boldsymbol{\theta}_v} J(\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_V), \tag{15}$$

其中，$\lambda_{\boldsymbol{\theta}_v}$ 是更新率，$\nabla_{\boldsymbol{\theta}_v} J(\boldsymbol{\theta}_1,\ \ldots\ldots,\ \boldsymbol{\theta}_V)$ 是全局 Q 值函数的梯度，其表达式为

$$\begin{aligned}\nabla_{\boldsymbol{\theta}_v} J(\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_V) &= \nabla_{\boldsymbol{\theta}_v}\left[Q_{tot}\left(\boldsymbol{s}^d,\boldsymbol{\alpha}^d\right) - r\left(\boldsymbol{s}^d,\boldsymbol{\alpha}^d\right) - \max_{\boldsymbol{\alpha}'} Q_{tot}(\boldsymbol{s}',\boldsymbol{\alpha}')\right]^2 \\ &= 2\Delta Q_{tot}\nabla_{\boldsymbol{\theta}_v} Q_{tot}\left(\boldsymbol{s}^d,\boldsymbol{\alpha}^d\right)\cdot\nabla_{\boldsymbol{\theta}_v} Q_{\boldsymbol{\theta}_v}\left(\boldsymbol{s}_v^d,\boldsymbol{\alpha}_v^d\right),\end{aligned} \tag{16}$$

其中，$\Delta Q_{tot} = Q_{tot}\left(\boldsymbol{s}^d,\ \boldsymbol{\alpha}^d\right) - r\left(\boldsymbol{s}^d,\ \boldsymbol{\alpha}^d\right) - \max_{\boldsymbol{\alpha}'} Q_{tot}(\boldsymbol{s}',\ \boldsymbol{\alpha}')$。结合熵最大化强化学习，[36], 策略 $\boldsymbol{\pi}_{\boldsymbol{\phi}_v}$ 的目标是期望奖励与动作熵的加权和。因此，策略网络的损失函数表示为

$$\begin{aligned}J_{\boldsymbol{\pi}}(\boldsymbol{\phi}_v) &= -\delta\mathcal{H}\left[\boldsymbol{\pi}_{\boldsymbol{\phi}_v}\left(\boldsymbol{\alpha}_v \mid \boldsymbol{s}_v^d\right)\right] - \mathbb{E}_{\boldsymbol{s}_v^d\in\mathcal{D},\boldsymbol{\alpha}_v\in\boldsymbol{\pi}_{\boldsymbol{\phi}_v}}\left[r_v\left(\boldsymbol{s}_v^d,\boldsymbol{\alpha}_v\right)\right] \\ &= \mathbb{E}_{\boldsymbol{s}_v^d\in\mathcal{D},\boldsymbol{\alpha}_v\in\boldsymbol{\pi}_{\boldsymbol{\phi}_v}}\left[\delta log\boldsymbol{\pi}_{\boldsymbol{\phi}_v}\left(\boldsymbol{\alpha}_v \mid \boldsymbol{s}_v^d\right) - Q_{\boldsymbol{\theta}_v}\left(\boldsymbol{s}_v^d,\boldsymbol{\alpha}_v\right)\right] \\ &= \mathbb{E}_{\boldsymbol{s}_v^d\in\mathcal{D}}\left[D_{KL}\left[\delta\boldsymbol{\pi}_{\boldsymbol{\phi}_v}\left(\boldsymbol{\alpha}_v \mid \boldsymbol{s}_v^d\right) \| \exp\left(Q_{\boldsymbol{\theta}_v}\left(\boldsymbol{s}_v^d,\boldsymbol{\alpha}_v\right)\right)\right]\right],\end{aligned} \tag{17}$$

其中，δ 是用于调整熵项权重的温度参数，随着 δ 增大，每个服务器的策略会更加随机。从（17）可以看出，

策略网络的目标等价于最小化条件概率分布 $\boldsymbol{\pi}_{\boldsymbol{\phi}_v}\left(\boldsymbol{\alpha}_v \mid \boldsymbol{s}_v^d\right)$ 与相应训练得到的个体Q值函数 $Q_{\boldsymbol{\theta}_v}\left(\boldsymbol{\alpha}_v \mid \boldsymbol{s}_v^d\right)$之间的Kullback-Leibler（KL）散度。因此，在给定局部观测的情况下，个体Q值较高的动作将获得较高的选择概率。然而，引入熵最大化后，其他具有潜在价值但选择概率较低的动作也能得到适当选择。因此，每个服务器探索有价值动作的能力以及找到最优资源块分配方案的概率都得到提升。最后，策略$\boldsymbol{\pi}_{\boldsymbol{\phi}_v}$ 可以通过梯度下降法按如下方式更新:

$$\boldsymbol{\phi}_v \leftarrow \boldsymbol{\phi}_v - \lambda_{\boldsymbol{\phi}_v} \nabla J_{\boldsymbol{\pi}}\left(\boldsymbol{\phi}_v\right), \tag{18}$$

其中， $\lambda_{\boldsymbol{\phi}_v}$ 为学习率。所提出的 VD-ERL算法的具体训练过程总结于**算法2**中。

C. 所提算法的复杂度与收敛性

本节分析所提出的用于面向语义的资源块分配的 VD-ERL算法的复杂度和收敛性。VD-ERL算法的复杂度主要来自语义三元组选择以及确定每个服务器的资源分配。首先，根据**算法1**，用户 k 的语义三元组选择复杂度为 $\mathcal{O}(N_{vk})$。因此，所有用户的语义三元组选择复杂度为 $\mathcal{O}\left(\sum_{v\in\mathcal{V}}\sum_{k=1}^{|\mathcal{U}_v|} N_{vk}\right) = \mathcal{O}(N)$。接着，我们说明每个服务器的策略网络和个体Q值网络的训练复杂度。这两个网络均为全连接网络，由输入层、隐藏层和输出层组成。因此，每个服务器网络训练的时间复杂度取决于各层神经元的数量 [37]。每个网络的时间复杂度为 $\mathcal{O}\left(\sum_{i=1}^{I-1} w_i w_{i+1} + (|\mathcal{L}_v|N_{vk}+Q)\ w_1 + |\mathcal{A}_v| w_I\right)$，其中 w_i 为隐藏层 i中的神经元数量， I 为隐藏层数， $|\mathcal{L}_v|N_{vk}+Q$ 和 $|\mathcal{A}_v|$ 分别表示输入层和输出层的维度。所提出的算法可以离线训练。因此，在训练阶段，**算法1** 和 **算法2** 的执行复杂度为 $\mathcal{O}\left(N + \sum_{i=1}^{I-1} w_i w_{i+1} + (|\mathcal{L}_v|N_{vk}+Q)\ w_1 + |\mathcal{A}_v| w_I\right)$。训练完成后，我们只需执行**算法1** 进行资源块分配，其复杂度为 $\mathcal{O}(N)$。

接下来，利用 [36, 定理 1] ,的结果，我们可以证明所提出的VD-ERL算法保证收敛到问题（8）的局部最优解，如下文所示

表 II：系统参数

参数	数值	参数	数值
Q	8	W	2 兆赫
V	5	U	50
P	1 W	N_0	-174 分贝毫瓦每赫兹
η	3	ρ	-3
ϵ	0.9	ξ	0.5

引理。

引理 1. 若满足以下条件，则所提出的 VD-ERL 算法必然收敛：1）个体Q值函数 $Q_{\boldsymbol{\theta_v}}(\boldsymbol{s}_v, \boldsymbol{\alpha}_v)$ 有界。2）对于任意状态 $\boldsymbol{s}_v$ 和动作 $\boldsymbol{\alpha}_v$，$Q^{\boldsymbol{\pi}^{\mathrm{N}}_{\phi_v}}_{\boldsymbol{\theta_v}}(\boldsymbol{s}_v, \boldsymbol{\alpha}_v) \geqslant Q^{\boldsymbol{\pi}^{\mathrm{O}}_{\phi_v}}_{\boldsymbol{\theta_v}}(\boldsymbol{s}_v, \boldsymbol{\alpha}_v)$ 成立，其中 $\boldsymbol{\pi}^{\mathrm{N}}_{\boldsymbol{\phi}_v}$ 是每次迭代中基于当前个体Q值函数、根据式（17）得到的优化策略。

证明： 接下来，我们证明所提出的 VD-ERL 算法满足这两个条件。由于所提出的 VD-ERL 算法中的动作数量 $\boldsymbol{\alpha}$ 有限，因此可以利用 [36, 引理 1] 证明全局Q值函数 $Q_{tot}(\boldsymbol{s}, \boldsymbol{\alpha})$ 有界。因此，条件 1）成立。对于条件 2），根据式（17），新策略 $\boldsymbol{\pi}^{\mathrm{N}}_{\boldsymbol{\phi}_v}$ 对任意旧策略 $\boldsymbol{\pi}^{\mathrm{O}}_{\boldsymbol{\phi}_v}$ 均满足以下不等式：

$$D_{KL}\left[\boldsymbol{\pi}^{\mathrm{N}}_{\boldsymbol{\phi}_v}(\boldsymbol{\alpha}_v \mid \boldsymbol{s}_v) \,\|\, \exp\left(Q_{\boldsymbol{\theta}_v}(\boldsymbol{s}_v, \boldsymbol{\alpha}_v)\right)\right] \leqslant D_{KL}\left[\boldsymbol{\pi}^{\mathrm{O}}_{\boldsymbol{\phi}_v}(\boldsymbol{\alpha}_v \mid \boldsymbol{s}_v) \,\|\, \exp\left(Q_{\boldsymbol{\theta}_v}(\boldsymbol{s}_v, \boldsymbol{\alpha}_v)\right)\right]. \quad (19)$$

根据（19），利用 [36,引理 2]的结果，可以证明所提出的方法满足条件2）。

IV. 仿真结果与分析

在仿真中，我们考虑一个圆形无线网络区域。在所考虑的网络中，五台服务器部署在中心周围，向均匀分布的 $U = 50$ 个用户传输图像数据。其他系统参数列于表 II。我们使用 [32] 中的场景图生成模型进行语义信息提取，并使用 [38] 中的多模态数据嵌入模型对语义信息和图像进行向量化。我们使用视觉基因组（VG） [39] 数据集训练所提出的算法。为进行比较，我们考虑了三种资源块分配方法作为基线：

算法2 用于求解问题 (8) 的 VD-ERL算法。

初始化： 网络参数 $\{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_V\}, \{\phi_1, \ldots, \phi_V\}$ 随机生成，学习率和更新率 $\{\lambda_{\boldsymbol{\theta}_1}, \ldots, \lambda_{\boldsymbol{\theta}_V}\}, \{\lambda_{\phi_1}, \ldots, \lambda_{\phi_V}\}$，以及迭代次数 N。

对于 $i = 1 \rightarrow N$ **执行**

  对于 每个环境步**执行**

    对于 每个智能体**执行**

      记录环境状态的局部观测 $\boldsymbol{s}_v$.

      根据当前策略 $\boldsymbol{\pi}_{\phi_v}$选择动作 $\boldsymbol{\alpha}_v$

      将动作 $\boldsymbol{\alpha}_v$ 和状态 $\boldsymbol{s}_v$ 传输给其他智能体。

      计算每台服务器的服务器奖励，并收集一系列轨迹 $\mathcal{D}_v = \mathcal{D}_v \cup \{(\boldsymbol{\alpha}_v, \boldsymbol{s}_v, r_v(\boldsymbol{s}_v, \boldsymbol{\alpha}_v))\}$。

    结束循环

  结束循环

  对于每个梯度步**执行**

    计算总奖励 $r(\boldsymbol{s}, \boldsymbol{\alpha})$ 和全局 Q 值 $Q_{tot}(\boldsymbol{s}, \boldsymbol{\alpha})$

    根据 (15) 更新 $\{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_V\}$。

    根据 (18) 更新 $\{\lambda_{\phi_1}, \ldots, \lambda_{\phi_V}\}$。

  结束循环

结束循环

资源块分配方法：a）随机方法，b）独立深度 Q 学习方法，以及 c）基于价值分解的深度 Q 学习网络方法。所有实验结果均取大量独立运行的平均值。

图 5 展示了一个使用我们设计的语义通信框架进行图像传输的示例。在图 5 中，服务器需要向用户发送一幅图像，如图 5a) 所示。然后，服务器使用场景图生成模型提取该图像的语义信息，如图 5b) 所示。在图 5b) 中，我们可以看到，原始图像（图 5a)）中的物体及其对应的关系均被提取出来。根据用户关联和资源块分配方案，下一步是选择三元组，以生成待传输的部分语义信息。图 5c) 展示了选出的三元组和生成的部分语义信息。从图 5b) 和图 5c) 可以看出，三元组“马旁边的路障”和三元组“马后面的树”未被选中用于生成语义信息，因为这些三元组无关紧要或存在冗余。这表明，所提出的图像语义通信框架能够找出无意义

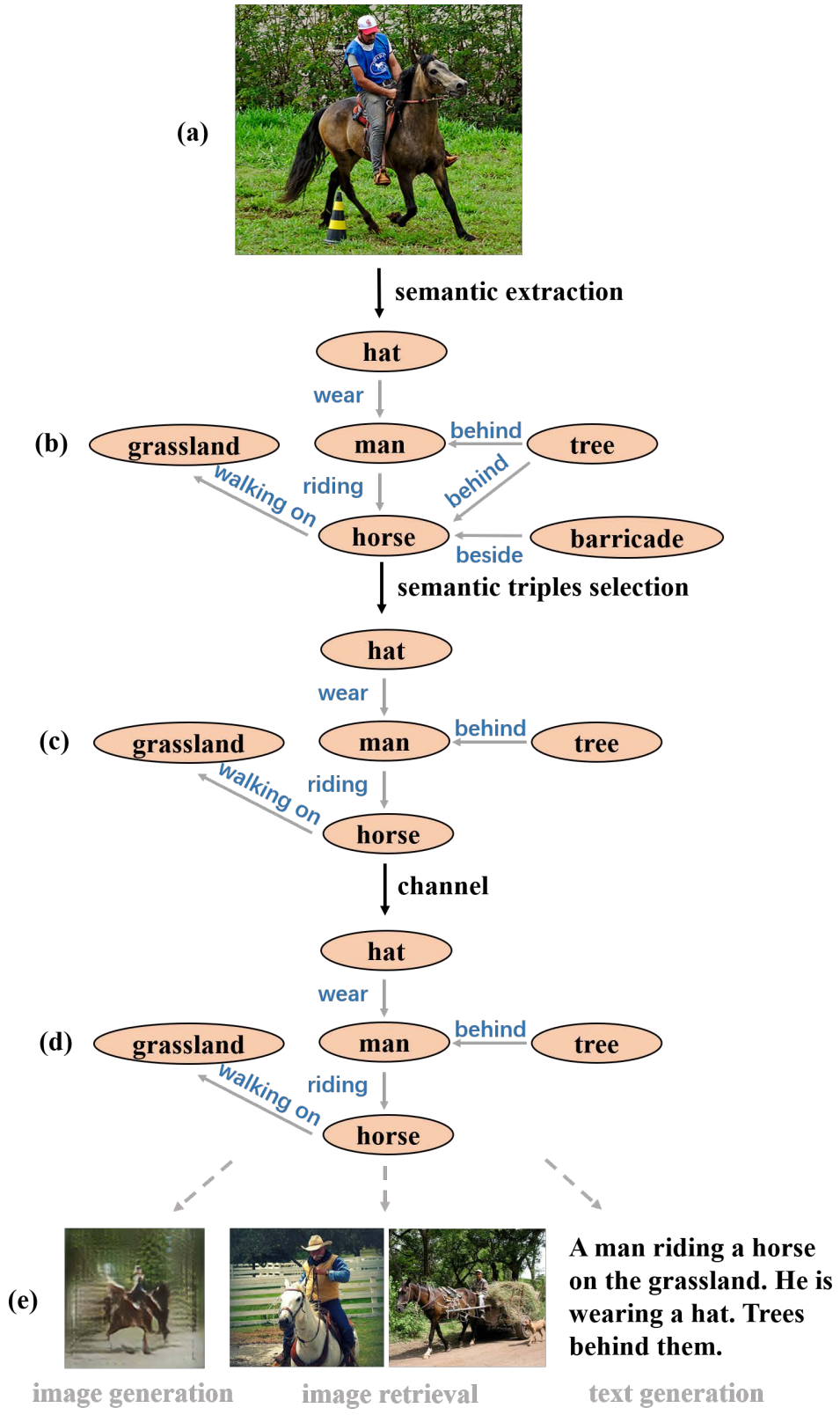


图 5. 图像传输语义通信系统示例。

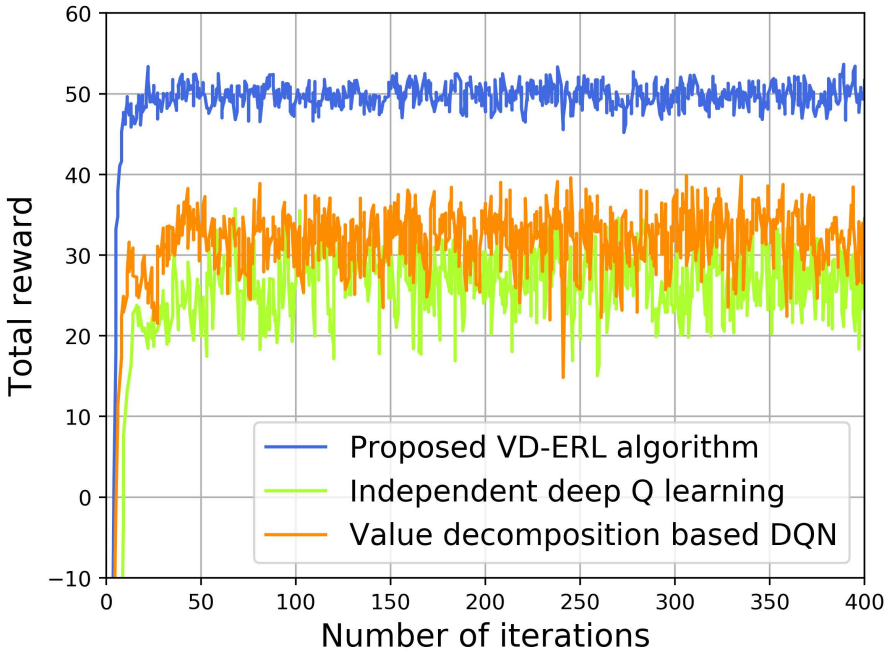
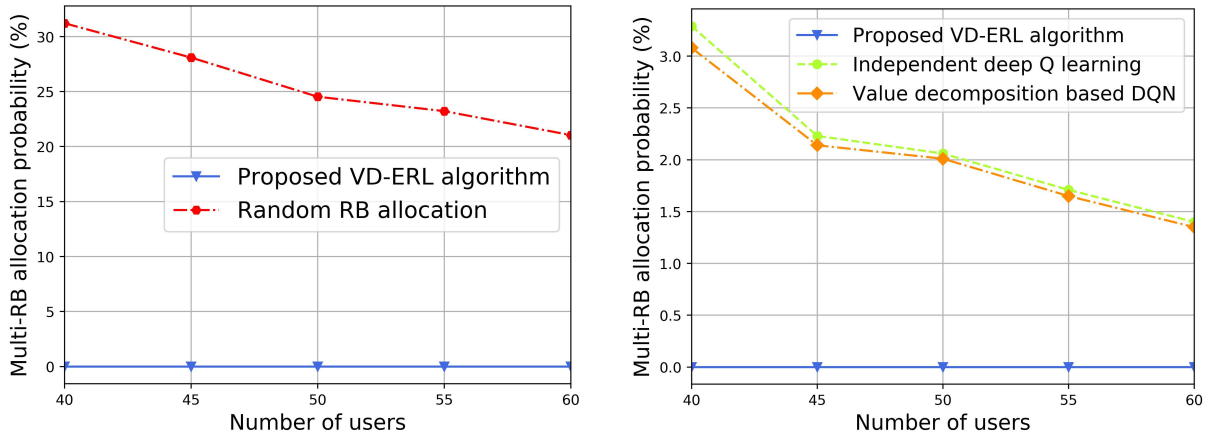


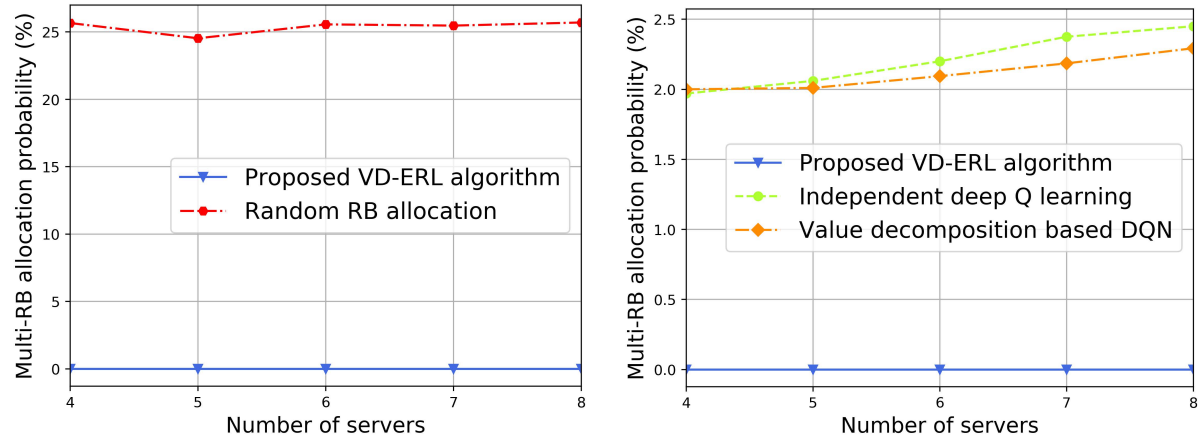
图 6. 所提出的 VD-ERL 算法的收敛情况。

三元组，并且不将其用于生成语义信息，从而仅通过传输部分重要三元组来降低传输时延。图 5d) 展示了用户接收到的语义信息。用户可以利用这些语义信息生成原始图像、检索相似图像，并生成原始图像的描述，如图 5e) 所示。具体而言，图 5e) 展示了如何利用生成对抗网络和接收到的语义信息，生成在语义层面上与原始图像相似的图像，这表明提取的语义信息足以支持各种应用。

图 6 展示了所提出的 VD-ERL算法的收敛情况。从图 6 可以看出，独立深度Q学习算法在经过 100 次迭代后仍未收敛。图 6 还表明，与经过 60 次迭代后收敛的基于VD的DQN算法相比，所提出的 VD-ERL算法在经过 30 次迭代后即可收敛。这是因为所提出的VD-ERL算法利用价值网络对策略网络进行评估和优化，因此价值函数的微小变化不会直接改变动作选择，而基于VD的DQN算法则不具备这一特性。从图 6 中还可以观察到，所提出的



(a) 用户数量变化时的随机资源块分配方法 (b) 用户数量变化时的传统多智能体强化学习方法。

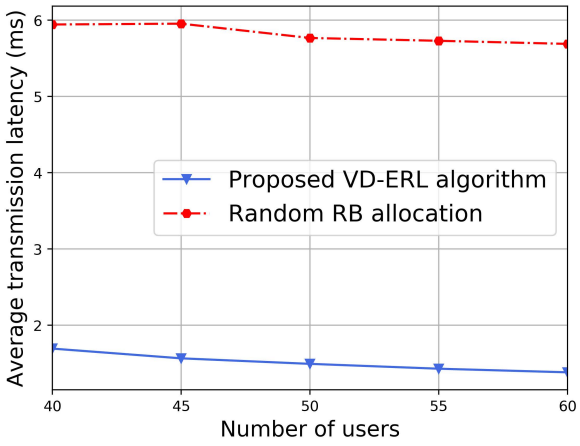


(c) 服务器数量变化时的随机资源块分配方法 (d) 服务器数量变化时的传统多智能体强化学习方法。

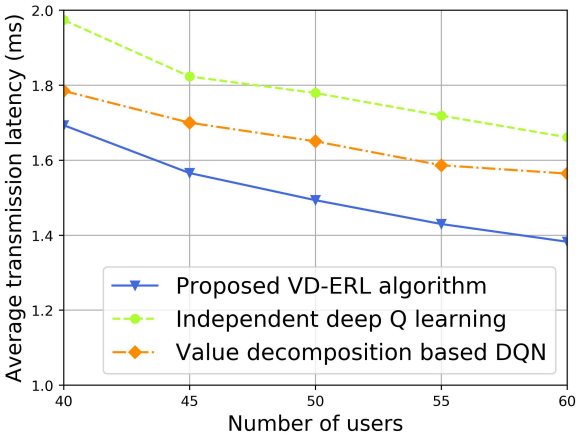
图 7. 所提出的 VD-ERL算法的多RB分配概率。

VD-ERL算法的总奖励相比独立深度Q学习算法和基于VD的DQN算法分别提升了 78.6% 和 42.9%。这是因为所提出的 VD-ERL能够通过最大化策略熵来优化动作探索，使每个服务器都能找到全局最优的资源块分配策略。

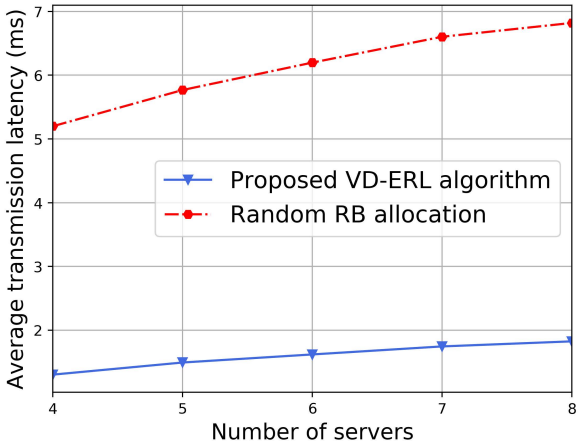
图 7 展示了多个服务器向一个用户分配资源块，而该用户仅使用来自一个服务器的一个资源块的概率如何分别随用户数量和服务器数量变化。下文将多个服务器向

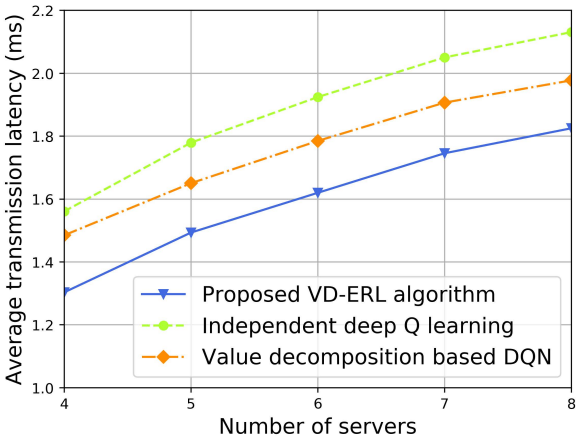


(a) 随用户数量变化与随机方法的比较。



(b) 随用户数量变化与传统的多智能体强化学习方法的比较。





(c) 随服务器数量变化与随机方法的比较。(d) 随服务器数量变化与传统的多智能体强化学习方法的比较。

图8. 所提出的 VD-ERL算法的平均传输时延。

一个用户分配资源块的概率定义为多RB分配概率。从图 7a)-7d) 可以看出，所提出的 VD-ERL算法产生的多RB分配概率为 0%，显著优于传统的多智能体强化学习算法。这是因为，所提出的旨在最大化期望总奖励的 VD-ERL算法能够使每个服务器与其他服务器协作，共同确定面向每个用户的资源块分配，从而避免多资源块分配。

图8展示了所有用户的平均传输时延随用户数量和服务器数量变化的情况。在图8a)和图8b)中可以看到，随着用户数量增加，所有所考虑算法的平均传输时延均有所下降。这是因为服务器能够利用有限的无线资源为 ISS 较高的用户提供服务。在图8c)和图8d)中可以看到，随着服务器数量增加，所有所考虑算法的平均传输时延均有所上升。这是因为服务器数量增加会加剧用户间的干扰，从而降低语义信息传输的数据速率。从图8中还可以观察到，与基线 a)、b) 和 c) 相比，所提出的 VD-ERL算法可分别将平均传输时延降低至多 74.1%、16.1% 和 9.5%。这是因为熵最大化与基于VD的深度强化学习框架相结合，使服务器能够协同探索资源块分配策略，从而最小化传输时延。

图9展示了语义三元组的语义分数分布与原始图像之间的关系。具体而言，随着语义分数增加，该语义三元组的颜色会从白色变为绿色。例如，图9b)中颜色最绿的语义三元组“女人拿着球拍”的语义分数为0.1351。从图9可以看出，语义分数较高的语义三元组更为重要，例如图9a)中的“男人骑滑板”和图9b)中的“女人拿着球拍”。图9还表明，根据语义分数排序，基于**算法1**，冗余语义三元组（如“男人穿着衬衫”）和不合理的语义三元组（如“男人站在树上”）的传输优先级低于其他三元组。

图10展示了所提出的VD-ERL算法与各基线算法的传输语义信息和资源块分配策略之间的相关性。具体而言，在图10a)中，随着语义得分增加，语义三元组的颜色从白色变为绿色。同样，在图10b)中，随着传输语义信息的ISS增加，传输的部分语义三元组的颜色也从白色变为绿色。从图10b)可以看出，随着传输的语义三元组数量增加，ISS单调递增。从图10a)和10b)可以看出，语义得分能够以较小误差描述每个语义三元组的重要性。例如，传输给用户5的语义信息的语义得分低于其他语义信息，其ISS也低于其他语义信息。图10c)、10d)和10e)分别展示了基线



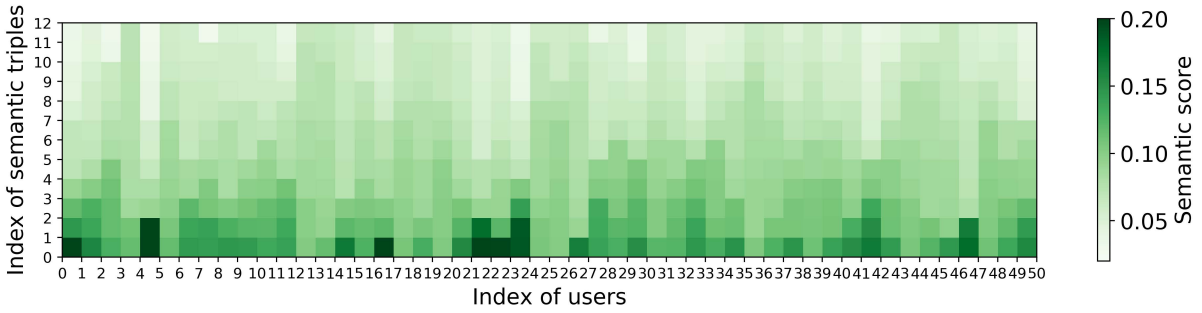
(a) 示例1



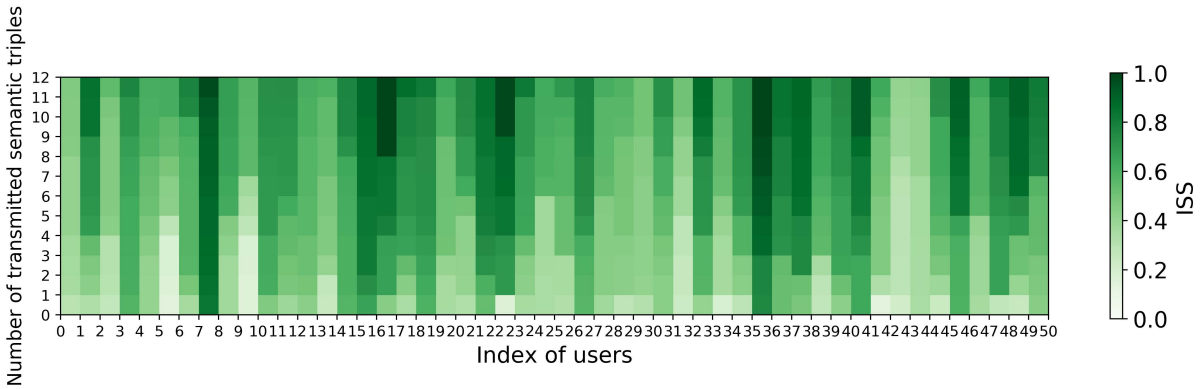
(b) 示例2

图9. 语义三元组的语义得分分布

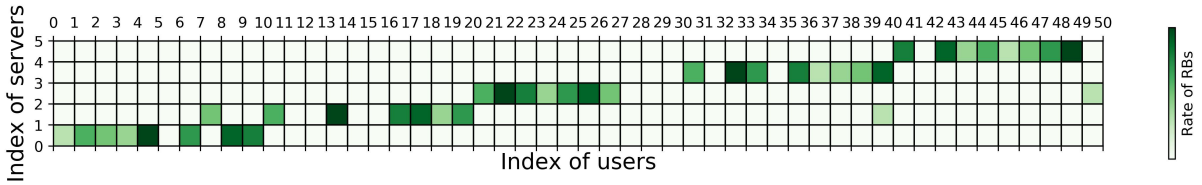
b)、基线c)和所提出的VD-ERL算法的资源块分配结果。具体而言，用户索引由用户与附近服务器之间的距离决定。距离最小的用户具有最小的索引。例如，用户0至9靠近服务器1，用户10至19靠近服务器2，用户20至29靠近服务器3，依此类推。在这些图中，随着资源块速率增加，该资源块的颜色会变得更绿。接着，从图10c)、10d)和10e)可以看出，与独立深度Q学习算法相比，所提出的VD-ERL算法使所有服务器能够协同为每个用户确定资源块分配。例如，如图10c)所示，服务器2和服务器4都打算为用户



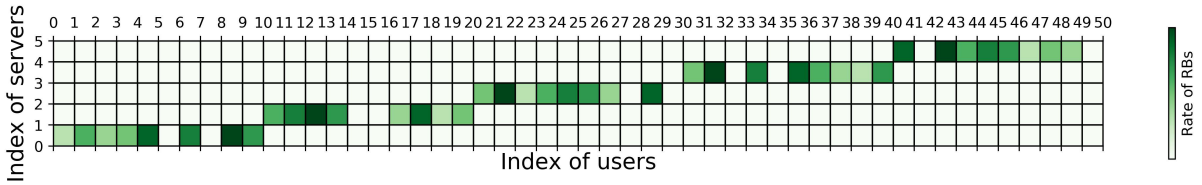
(a) 用户语义得分分布



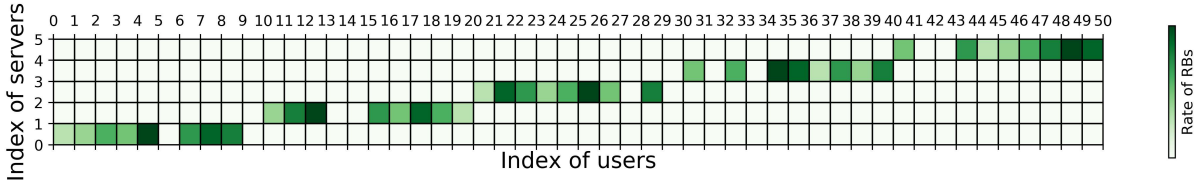
(b) ISS 随用户传输的语义三元组数量变化



(c) 基于基线 b) 的资源块分配



(d) 基于基线 c) 的资源块分配



(e) 基于 VD-ERL算法的资源块分配

图 10. 传输的语义信息与资源块分配策略之间的相关性

39，这会导致一个多RB分配问题。

V. 结论

本文提出了一种新颖的图像语义通信框架，使一组服务器能够利用语义通信技术协作地向各自关联的用户传输图像。我们将每幅图像的语义信息建模为场景图，其中包含一组物体及其相互关系。我们提出了一种ISS指标，用于评估原始图像与其文本语义信息之间的语义相似度。在无线资源受限的情况下，每台服务器都必须联合确定要传输的语义信息和RB分配方案。我们将该问题表述为一个优化问题，目标是在满足ISS要求的同时，最小化平均传输时延。为解决该问题，我们开发了一种基于价值分解的熵最大化多智能体强化学习算法，使服务器能够基于各自的局部观测，找到最优的协作式RB分配方案。仿真结果表明，与传统的多智能体强化学习算法相比，所提出的算法显著降低了传输时延，并提高了收敛速度。

参考文献

[1] W. Zhang, Y. Wang, M. Chen, T. Luo, and D. Niyato, 《语义通信网络中的图像传输优化》，载于*IEEE*国际全球通信会议论文集，巴西里约热内卢，2022年12月。[2] M. Chen, D. Gündüz, K. Huang, W. Saad, M. Bennis, A. V. Feljan, and H. V. Poor, 《无线网络中的分布式学习：最新进展与未来挑战》，*IEEE*通信领域精选期刊，第39卷，第12期，2021年12月。[3] X. Mu, Y. Liu, L. Guo, and N. Al-Dhahir, 《异构语义与比特通信：一种半NOMA方案》，在线发布：https://arxiv.org/abs/2205.02620，2022年。[4] G. Zhu, D. Liu, Y. Du, C. You, J. Zhang, and K. Huang, 《迈向智能边缘：无线通信与机器学习相遇》，*IEEE CommunicationsMagazine*，第58卷，第1期，第19–25页，2020年。[5] C. Chaccour, W. Saad, M. Debbah, Z. Han, and H. V. Poor, 《更少的数据，更多的知识：构建下一代语义通信网络》，在线发布：https://arxiv.org/abs/2211.14343，2022年。[6] M. Kalfa, S. Y. Yetim, A. Atalik, M. Gok, Y. Ge, R. Li, W. Tong, T. M. Duman, and O. Arikan, 《图信号语义信息的可靠提取与创新率估计》，在线发布：https://arxiv.org/abs/2211. 05440，2022年。[7] J. Kang, H. Du, X. Li, Z. Xiong, S. Ma, D. Niyato, and Y. Li, 《面向任务的语义通信中的个性化显著性：图像传输与性能分析》，在线发布：https://arxiv.org/abs/2209.12274，2022年。

[8] Z Qin, X. Tao, J. Lu, and G. Y. Li, 《语义通信：原理与挑战》，在线发布：https: // arxiv.org/abs/2201.01389，2022年。[9] X. Luo, H. Chen, and Q. Guo, 《语义通信：概述、开放问题与未来研究方向》，*IEEEWireless Communications*，第29卷，第1期，第210–219页，2022年1月。[10] Z. Lin, Y. Gong, and K. Huang, 《用于快速分布式优化的分布式空中计算：波束成形设计与收敛性分析》，在线发布：https://arxiv.org/abs/2204.06876v1，2022年。[11] H. Zou, C. Zhang, S. Lasaulce, L. Saludjian, and H. V. Poor, 《目标导向量化：分析、设计及其在资源分配中的应用》，在线发布：https://arxiv.org/abs/2209.15347，2022年。[12] W. Saad, M. Bennis, and M. Chen, 《6G无线系统展望：应用、趋势、技术与开放研究问题》，*IEEE Network*，第34卷，第3期，第134–142页，2020年5月。[13] P. Tandon, S. Chandak, P. Pataranutaporn, Y. Liu, A. M. Mapurange, P. Maes, T. Weissman, and M. Sra, 《Txt2vid：通过文本实现人头说话视频的超低比特率压缩》，在线发布：https://arxiv.org/abs/2106.14014v3，2022年。[14] Z. Meng, C. She, G. Zhao, and D. D. Martini, 《用于元宇宙中现实世界设备与数字模型同步的采样、通信与预测协同设计》，在线发布：https://arxiv.org/abs/2208.04233，2022年。[15] G. Shi, Y. Xiao, Y. Li, and X. Xie, 《从语义通信到语义感知网络：模型、架构与开放问题》，*IEEE CommunicationsMagazine*，第59卷，第8期，第44–50页，2021年9月。[16] T. Han, Q. Yang, Z. Shi, S. He, and Z. Zhang, 《用于高效语音传输的语义保持通信系统》，在线发布：https://arxiv.org/abs/2205.12727，2022年。[17] P. Jiang, C. K. Wen, S. Jin, and G. Y. Li, 《用于视频会议的无线语义通信》，在线发布：https://arxiv.org/pdf/2204.07790v1，2022年。[18] J. Bao, P. Basu, M. Dean, C. Partridge, A. Swami, W. Leland, and J. A. Hendler, 《迈向语义通信理论》，载于*IEEE*网络科学研讨会论文集，美国纽约州西点，2011年6月。[19] P. Basu, J. Bao, M. Dean, and J. A. Hendler, 《利用语义关系保持信息质量》，载于*IEEE*普适计算与通信国际会议研讨会论文集，瑞士卢加诺，2012年3月。[20] H. Zhang, S. Shao, M. Tao, X. Bi, and K. B. Letaief, 《基于深度学习的语义通信系统：任务无感知发射机与动态数据》，在线发布：https://arxiv.org/abs/2205.00271，2022年。[21] Y. Yang, C. Guo, F. Liu, C. Liu, L. Sun, Q. Sun, and J. Chen, 《面向人工智能任务的语义通信》，在线发布：https://arxiv.org/abs/2109.14170，2021年。[22] Y. Wang, M. Chen, T. Luo, W. Saad, D. Niyato, H. V. Poor, and S. Cui, 《语义通信性能优化：一种基于注意力的强化学习方法》，*IEEE*通信领域精选期刊，第2598–2613页，2022年7月。[23] K. Lu, R. Li, X. Chen, Z. Zhao, and H. Zhang, 《通过语义相似度实现强化学习赋能的语义通信》，在线发布：https://arxiv.org/abs/2108.12121，2021年8月。[24] H. Xie, Z. Qin, and G. Y. Li, 《面向视觉问答的面向任务的多用户语义通信》，*IEEE WirelessCommunications Letters*，第11卷，第3期，第553–557页，2022年12月。[25] J. Wang, Y. Duan, X. Tao, M. Xu, and J. Lu, 《基于卷积网络拉普拉斯金字塔的语义感知图像压缩》，*IEEE*图像处理汇刊，第30卷，第4225–4237页，2021年3月。[26] D. Huang, F. Gao, X. Tao, Q. Du, and J. Lu, 《迈向语义通信：基于深度学习的图像语义编码》，在线发布：https://arxiv.org/abs/2208.04094，2022年8月。

[27] X. Li, J. Shi, and Z. Chen, “通过强化学习实现任务驱动的语义编码”，*IEEE*图像处理汇刊，第30卷，6307–6320页，2021年7月。[28] M. Chen, Y. Wang, and H. V. Poor, “能量采集网络中无线语义通信的性能优化”，载于*IEEE*声学、语音与信号处理国际会议论文集，新加坡，新加坡，2022年5月。[29] A. Tampuu, T. Matiisen, D. Kodelja, I. Kuzovkin, K. Korjus, J. Aru, J. Aru, and R. Vicente, “深度强化学习下的多智能体合作与竞争”，在线获取：http://arxiv.org/abs/1511.08779，2015年11月。[30] K. Tang, H. Zhang, B. Wu, W. Luo, and W. Liu, “学习构建用于视觉上下文的动态树结构”，载于*IEEE/CVF*计算机视觉与模式识别会议论文集，美国加利福尼亚州长滩，2019年6月。[31] S. Lee, J. Kim, Y. Oh, and J. H. Jeon, “基于场景图的视觉问答”，载于第一届图计算国际会议论文集，美国加利福尼亚州拉古纳希尔斯，2019年9月。[32] K. Tang, Y. Niu, J. Huang, J. Shi, and H. Zhang, “从偏置训练中无偏地生成场景图”，载于*IEEE/CVF*计算机视觉与模式识别会议论文集，美国华盛顿州西雅图，2020年6月。[33] W. Zhou, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “图像质量评估：从误差可见性到结构相似性”，*IEEE*图像处理汇刊，第13卷，第4期，600–612页，2004年4月。[34] M. Corbetta and G. L. Shulman, “大脑中目标导向注意与刺激驱动注意的控制”，《自然综述：神经科学》，第3卷，第3期，201–215页，2002年3月。[35] P. Sunehag, G. Lever, A. Gruslys, W. M. Czarnecki, V. Zambaldi, M. Jaderberg, M. Lanctot, N. Sonnerat, J. Z. Leibo, K. Tuyls, and T. Graepel, “用于合作式多智能体学习的价值分解网络”，在线获取：http: //arxiv.org/abs/1706.05296，2017年6月。[36] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, “软演员-评论家：采用随机演员的离策略最大熵深度强化学习”，在线获取：http://arxiv.org/abs/1801.01290，2018年1月。[37] Y. Wang, M. Chen, Z. Yang, T. Luo, and W. Saad, “利用深度学习优化配备可见光通信的无人机部署”，*IEEE*无线通信汇刊，第19卷，第11期，7049–7063页，2020年11月。[38] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever, “从自然语言监督中学习可迁移视觉模型”，在线获取：https: //arxiv.org/abs/2103.00020，2021年2月。[39] R. Krishna, Y. Zhu, O. Groth, J. Johnson, K. Hata, J. Kravitz, S. Chen, Y. Kalantidis, L. Li, D. A. Shamma, M. S. Bernstein, and F. Li, “视觉基因组：利用众包密集图像标注连接语言与视觉”，在线获取：https://arxiv.org/abs/1602.07332，2016年2月。