

Comunicação Semântica e Aprendizado Federado: Uma Arquitetura baseada em Autoencoders para Redes 6G

Gian P. Rodrigues e Herman L. dos Santos

Resumo— Com a ascensão das redes 6G, a comunicação semântica emerge como um paradigma disruptivo para superar os limites de Shannon, ao focar na transmissão do significado da informação, em vez da reconstrução exata de bits. Este trabalho propõe e avalia uma arquitetura federada baseada em autoencoders para compressão semântica orientada a tarefas no banco de dados CIFAR-10. Por meio de experimentos exaustivos, demonstra-se que o uso de espaços latentes compactos integrados ao Aprendizado Federado permite reduções superiores a 97% no tráfego de rede, sem perda semântica. O *framework* proposto leva em consideração amostras ruidosas, geradas artificialmente durante o treinamento, atuando como um mecanismo de regularização e contribuindo para a robustez semântica, com ganhos expressivos de acurácia. Os resultados obtidos validam a eficácia de sistemas orientados a tarefas em ambientes altamente afetados por desvanecimento e com recursos restritos, estabelecendo um novo compromisso entre taxa e desempenho para o ecossistema de Inteligência Artificial distribuída.

Palavras-Chave— Comunicação Semântica, Aprendizado Federado, Orientação a Tarefas, Information Bottleneck.

I. INTRODUÇÃO

A arquitetura das redes de comunicação modernas está atingindo um ponto de inflexão fundamental. Por décadas, o design de sistemas seguiu rigorosamente a Teoria Matemática da Comunicação de Shannon [1], focada na transmissão fidedigna de símbolos. No entanto, com a explosão de dados gerados por inteligência artificial (IA) e a saturação do espectro, a busca por maiores taxas de transmissão de bits torna-se insuficiente para atender às demandas de latência e de volume de dados em redes futuras. O paradigma da sexta geração de telecomunicações (6G) propõe uma evolução necessária nos níveis *B* – Semântico e *C* – Eficácia, com o objetivo principal de assegurar o sucesso da tarefa final no receptor, descartando qualquer redundância estatística irrelevante para o significado da mensagem [2].

Este modelo, conhecido como comunicação semântica (SemCom), é particularmente vital para a IA generativa (GenAI) em ambientes distribuídos, onde a geração e o consumo de dados ocorrem nos nós clientes, frequentemente sob restrições severas de largura de banda e de processamento. Dispositivos móveis, veículos autônomos e sensores remotos frequentemente operam em condições de canal adversas e com

autonomia energética limitada. Transmitir grandes volumes de dados brutos sobrecarrega a rede. Como solução, a SemCom permite que esses dispositivos extraiam uma "essência semântica" compacta, transmitindo apenas o conhecimento estritamente necessário para que um receptor inteligente realize inferências ou reconstruções precisas, minimizando o uso de recursos de rádio.

Simultaneamente, o aprendizado federado (FL) permite o treinamento de modelos globais, preservando a privacidade dos dados locais, requisito obrigatório em aplicações críticas de saúde e segurança [3]. No entanto, o FL sofre com o gargalo de comunicação na agregação frequente de parâmetros entre o servidor central e os nós locais. Integrar SemCom ao FL cria uma sinergia poderosa em que os agentes não apenas compartilham pesos, mas também trocam representações latentes robustas. Este trabalho investiga como autoencoders convolucionais podem servir como codificadores semânticos leves, analisando detalhadamente o impacto das dimensões latentes e da condição de canal sobre a acurácia e a resiliência do sistema global.

As principais contribuições deste estudo são detalhadas a seguir:

- Propor uma arquitetura SemCom-FL multitarefa integrando classificação orientada a tarefas e reconstrução visual estrutural;
- Analisar quantitativamente o compromisso entre compressão extrema e retenção de acurácia discriminativa sob restrições severas de banda;
- Demonstração teórica e prática de que o ruído de canal atua como um regularizador intrínseco: sob a ótica do *information bottleneck*, o ruído força o codificador a reter apenas as informações essenciais para a tarefa, promovendo ganhos de generalização;
- Validação empírica no banco de dados CIFAR-10 [4], mostrando robustez superior em relação a métodos tradicionais de compressão de dados brutos e transmissões ponto-a-ponto.
- Discussão sobre as implicações de privacidade, latência e eficiência energética em redes 6G sustentáveis e IA-Nativas.

II. TRABALHOS RELACIONADOS

O campo da comunicação semântica tem sido revitalizado pelo avanço exponencial do aprendizado profundo. A SemCom pode ser caracterizada como o pilar das comunicações

centradas no conhecimento, em que o receptor interpreta pistas semânticas para reconstruir a intenção original do transmissor [2]. Luo et al. introduz o conceito de comunicação orientada a tarefas, demonstrando que a extração de características específicas para uma aplicação pode superar os limites tradicionais de Shannon em termos de eficiência no uso do espectro e de robustez a condições adversas do canal [5].

Frameworks consolidados, e.g. DeepSC, utilizam redes neurais profundas treinadas ponta a ponta para otimizar a transmissão de texto e imagem sobre canais ruidosos e desvanecidos [6]. Entretanto, a aplicação sistemática desses conceitos em FL ainda é uma fronteira aberta e desafiadora. O uso de GenIA para auxiliar a SemCom sugere que o receptor pode "imaginar" os detalhes perdidos desde que a base semântica seja transmitida de forma íntegra [7]. Este trabalho foca na implementação de codificadores convolucionais leves que possam ser efetivamente embarcados em dispositivos locais com baixa capacidade computacional.

III. METODOLOGIA PROPOSTA

A metodologia empregada organiza-se em um pipeline de processamento semântico distribuído, ilustrado na Figura 1. Conforme o diagrama ilustra, o dispositivo cliente captura os dados brutos e os codifica num espaço latente compacto utilizando uma rede neural convolucional (CNN). O canal sem fio é modelado por meio da injeção de ruído nesse vetor, que então é bifurcado em duas tarefas simultâneas: a classificação (objetivo principal) e a reconstrução (regularização). Por fim, o processo se integra ao aprendizado federado por meio do envio dos pesos atualizados de volta ao servidor global para agregação por *Federated Averaging* (FedAvg).

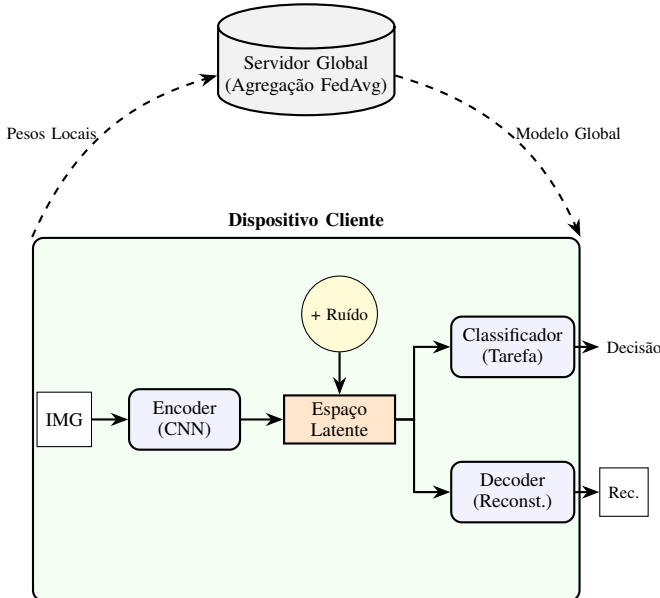


Fig. 1. Arquitetura detalhada da comunicação semântica federada multitarefa.

A. Codificação Semântica em Imagens no Cliente

Diferentemente de dados textuais, que são inerentemente discretos e exigem arquiteturas baseadas em atenção para

capturar dependências complexas de longo alcance, as imagens possuem alta redundância espacial e natureza contínua. Isso permite o uso de CNNs para extrair características hierárquicas por meio de operações de filtragem espacial. O Codificador $\mathcal{E}$ empregado neste trabalho utiliza camadas de convolução seguidas de *pooling* para reduzir a dimensionalidade espacial, mapeando o dado bruto $\mathbf{x}$ para um vetor latente compacto $\mathbf{z} \in \mathbb{R}^L$, em que L é a dimensão do vetor latente. Esta escolha técnica garante uma pegada computacional mínima localmente, permitindo que dispositivos com hardware limitado realizem compressão em tempo real, sem degradação excessiva da latência e/ou da eficiência energética.

B. Mecânica do Variational Autoencoder e Espaço Latente

A arquitetura baseia-se nos princípios dos *Variational Autoencoders* (VAEs), em que o Encoder projeta a entrada em uma distribuição probabilística no espaço latente. A escolha de L define o rigor do *semantic bottleneck* aplicado aos dados. Durante a transmissão entre cliente e servidor, as imperfeições no conhecimento de canal são modeladas através da injeção de ruído branco Gaussiano aditivo (AWGN) $n \sim \mathcal{N}(0, \sigma^2)$. A representação recebida no servidor apresentada na Eq. (1).

$$\tilde{\mathbf{z}} = \mathbf{z} + \mathbf{n}. \quad (1)$$

A injeção de ruído artificial é um processo crítico para a robustez do sistema, pois não apenas modela uma interferência indesejada, mas também atua como uma poderosa ferramenta de regularização estocástica. Ele força o codificador a criar vizinhanças semânticas robustas, garantindo que pequenas flutuações ou desvios no vetor latente não alterem drasticamente a interpretação final da tarefa pretendida.

C. Validação Task-Oriented via Classificador Latente

Para validar se a semântica essencial para a tomada de decisão é preservada, acopla-se um classificador latente $\mathcal{C}$ diretamente à saída do codificador. Esta abordagem *task-oriented* propõe que, mesmo sob compressão extrema, os vetores latentes contêm informação discriminativa suficiente para a classificação correta. O Classificador Latente atua como um "filtro de inteligência", descartando ruídos de pixels que não contribuem para a definição da classe, e funciona como uma compressão extrema que prioriza a *intenção* semântica sobre a *forma* bruta dos dados. O Decodificador $\mathcal{D}$ atua como uma tarefa secundária de reconstrução, auxiliando na regularização topológica do espaço latente ao forçar o modelo a manter relações estruturais com o espaço de pixels original. A perda multitarefa que guia o treinamento federado é definida na Eq. (2).

$$\mathcal{L} = \mathcal{L}_{\text{CE}}(\mathcal{C}(\tilde{\mathbf{z}}), \mathbf{y}) + \alpha \mathcal{L}_{\text{MSE}}(\mathcal{D}(\tilde{\mathbf{z}}), \mathbf{x}), \quad (2)$$

onde $\mathcal{L}_{\text{CE}}$ é a entropia cruzada para a tarefa de classificação e $\mathcal{L}_{\text{MSE}}$ é o erro quadrático médio para a tarefa de reconstrução, com o hiperparâmetro de balanço fixado em $\alpha = 0,5$.

D. Processo de Agregação FedAvg

A atualização do modelo global no servidor central segue o algoritmo clássico FedAvg. Após o treinamento local em cada k -ésimo dispositivo cliente, o servidor calcula a média ponderada dos parâmetros em cada t -ésima iteração através da Eq. (3),

$$w_{t+1} = \sum_{k=1}^K \frac{n_k}{n} w_t^k, \quad (3)$$

onde n_k é o número de amostras no cliente k e N é o total de amostras. A integração da SemCom reduz drasticamente o tempo de transmissão de w_t^k , acelerando a convergência global.

IV. CONFIGURAÇÃO EXPERIMENTAL

As avaliações foram conduzidas utilizando o banco de dados CIFAR-10 [4], que consiste em 60.000 imagens coloridas de 32×32 pixels, distribuídas em 10 classes mutuamente exclusivas. O cenário federado simulou três clientes com distribuição de dados independentes e identicamente distribuídos. O treinamento foi executado por três rodadas federadas, $T = 3$, com uma época local por rodada, utilizando o otimizador Adam com taxa de aprendizado fixada em 10^{-3} .

TABELA I
HIPERPARÂMETROS DE TREINAMENTO FEDERADO.

Parâmetro	Valor
Otimizador	Adam
Taxa de Aprendizado (LR)	1×10^{-3}
Batch Size (Local)	64
Rodadas Federadas (T)	3
Épocas Locais (E)	1
Dataset	CIFAR-10
Hiperparâmetro α	0,5
Dispositivos Clientes	3
Seed Aleatória	42

A. Arquitetura das Redes Neurais

O Encoder é composto por três blocos convolucionais com filtros $\{32, 64, 128\}$ e kernel 3×3 , seguidos de camadas de pooling e de uma camada densa final para o latente. O Decoder espelha essa estrutura utilizando convoluções transpostas para a reconstrução. O Classificador Latente consiste em uma rede densa de duas camadas com ativação ReLU e saída Softmax para as 10 classes. Esta configuração, obtida empiricamente, equilibra profundidade e custo computacional. Um exemplo de reconstrução é apresentado na Figura 2

V. RESULTADOS E DISCUSSÃO

A análise dos resultados revela um comportamento dinâmico e resiliente na interseção entre a compressão semântica e a robustez de canal. A Tabela II apresenta os dados comparativos essenciais obtidos.

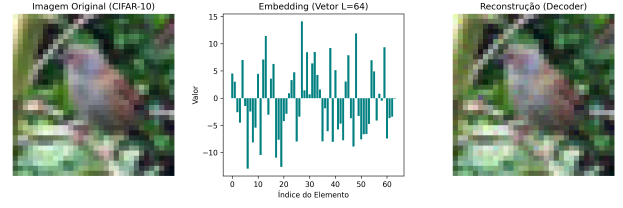


Fig. 2. Visualização de dados experimentais reais: (a) Imagem original do CIFAR-10, (b) Vetor de embedding e (c) Reconstrução via decoder.

TABELA II
RESULTADOS AGREGADOS: ACURÁCIA VS. CUSTO DE COMUNICAÇÃO.

Latent Dim	Noise	Acc. Base	Acc. Comp.	CR	Cost (bits)
–	0.00	0.6771	–	1x	4.92×10^9
16	0.00	–	0.6066	192x	2.56×10^7
16	0.05	–	0.6220	192x	2.56×10^7
16	0.10	–	0.6188	192x	2.56×10^7
32	0.00	–	0.6404	96x	5.12×10^7
32	0.05	–	0.6434	96x	5.12×10^7
32	0.10	–	0.6395	96x	5.12×10^7
64	0.00	–	0.6488	48x	1.02×10^8
64	0.05	–	0.6509	48x	1.02×10^8
64	0.10	–	0.6492	48x	1.02×10^8

A. Eficiência de Comunicação Semântica e Taxa-Acurácia

A transmissão integral de bits (linha de base) requer impressionantes $4,92 \times 10^9$ bits para atingir uma acurácia de 0,6771. Em contrapartida, a arquitetura semântica proposta com $L = 64$ atinge uma acurácia de 0,6509, consumindo apenas $1,02 \times 10^8$ bits. Isso representa uma economia de tráfego de 97,9% com uma penalidade de desempenho marginal de apenas 2,6 pontos percentuais. Para dimensões latentes ainda menores, como $L = 16$, a economia de banda atinge 99,48%, mantendo uma acurácia de 0,6220 sob ruído moderado, $\sigma = 0,05$, o que, nos experimentos, equivale a uma relação sinal-ruído (SNR) empírica de aproximadamente 26 dB no espaço latente. Este resultado é fundamental para o ecossistema 6G, pois sugere que uma grande parcela dos bits transmitidos em comunicações convencionais é redundância estatística, que pode ser abstraída sem comprometer o significado da mensagem.

B. O Fenômeno da Regularização por Ruído

Um dos achados mais disruptivos deste estudo é o impacto positivo do ruído moderado no processo de aprendizado semântico. Conforme ilustrado na Figura 3, a acurácia na configuração $L = 64$ com $\sigma = 0,05$ (0,6509) superou a do cenário ideal sem ruído (0,6488). Este comportamento, embora contraintuitivo para a engenharia de comunicações clássica, fundamenta-se na teoria de *information bottleneck* [8]. O ruído gaussiano atua como uma barreira que impede o codificador de memorizar detalhes irrelevantes e ruídos de pixel da entrada, o que causa *overfitting*. Em vez disso, o codificador é forçado a focar exclusivamente em invariantes estruturais e características semânticas globais, resilientes a flutuações de canal.

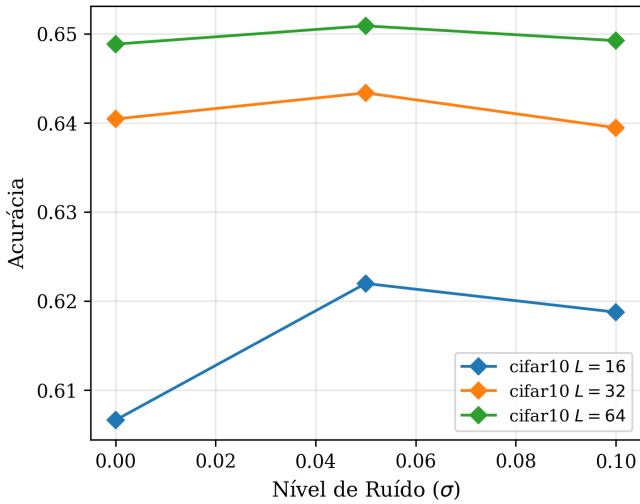


Fig. 3. Impacto do nível de ruído na acurácia do sistema para diferentes dimensões latentes. (Nota: Para $\sigma = 0,05$, a Relação Sinal-Ruído empírica é de aproximadamente 26 dB no espaço latente).

C. Hierarquia Semântica e Dimensão Latente

A Figura 4 demonstra que a queda de acurácia em função da compressão não é linear, e sim logarítmica. Isso sugere que o codificador semântico aprende a priorizar informações de forma hierárquica. Os primeiros bits do vetor latente capturam as características mais críticas para a distinção entre classes, *e.g.* formas globais de um "avião" vs "carro", enquanto os bits subsequentes adicionam detalhes de textura e refinamento. Esta propriedade permite que o sistema 6G ajuste dinamicamente a dimensão latente com base na qualidade de serviço exigida e nas condições de desvanecimento e conhecimento do canal, otimizando o uso do espectro em tempo real.

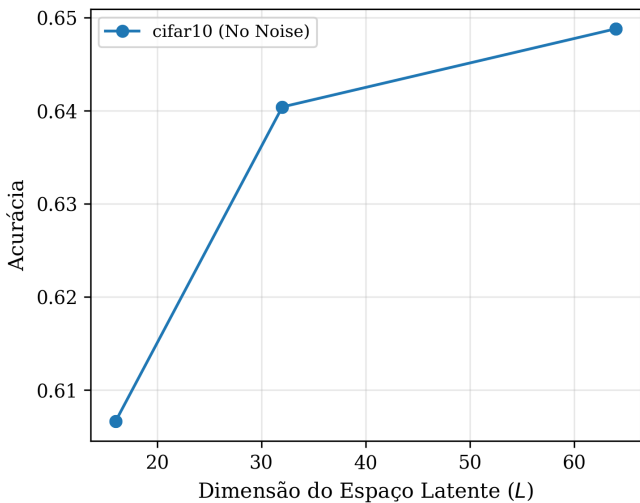


Fig. 4. Análise da acurácia em função da dimensionalidade do espaço latente.

VI. DISCUSSÃO TÉCNICA AVANÇADA

Nesta seção, exploram-se as implicações da arquitetura proposta sob as óticas de eficiência energética, robustez e

privacidade em redes de larga escala.

A. Cenários 6G e Casos de Uso: A arquitetura SemCom-FL proposta é ideal para cenários de cidades inteligentes, em que milhares de câmeras locais precisam realizar monitoramento de tráfego em tempo real, sem sobrecarregar o *backhaul* da rede. Em carros autônomos, a transmissão de essências semânticas permite uma coordenação coletiva rápida, mesmo sob interferência severa, priorizando a segurança em relação à fidelidade visual.

B. Eficiência Energética e Sustentabilidade: A redução drástica no volume de dados transmitidos, de até 192 vezes, tem impacto direto no consumo de energia dos terminais locais. Em dispositivos alimentados por bateria, o subsistema de comunicação é frequentemente o maior consumidor de energia. Ao processar semanticamente os dados localmente e transmitir apenas vetores latentes compactos, estende-se a vida útil operacional de sensores remotos e reduz-se a pegada de carbono da infraestrutura de rede, um dos pilares de sustentabilidade do 6G.

C. Privacidade e Segurança Natural: Uma vantagem secundária, porém crítica, da SemCom-FL é a privacidade. Ao transmitir apenas representações latentes compactas em vez de dados brutos ou gradientes de modelos completos, adiciona-se uma camada natural de ofuscação. Ataques de reconstrução dos dados originais tornam-se significativamente mais complexos quando o dispositivo *eavesdropper* tem acesso apenas a um vetor latente comprimido e ruidoso, o que reforça a segurança do ecossistema federado.

D. Escalabilidade Massiva e mMTC A escalabilidade é um desafio central para as redes de próxima geração. Ao reduzir o custo de comunicação em uma ordem de magnitude em termos de bits, o sistema proposto permite que um único ponto de acesso suporte uma ordem de magnitude maior de dispositivos simultâneos. Este ganho de eficiência é um requisito fundamental para viabilizar o cenário *ultra massive Machine Type Communications*, previsto para 6G, no qual a densidade de conexões exige que a rede gerencie recursos de forma ultra-parcimoniosa por meio de inteligência semântica.

E. Relação Sinal-Ruído Semântica (S-SNR): Diferentemente da SNR tradicional, que foca na potência física do sinal *versus* a potência de ruído, a abordagem semântica introduz o conceito de SNR semântica (S-SNR), focada na preservação da *utilidade* da informação para a tarefa. Os experimentos apresentados sugerem que a S-SNR permanece alta mesmo quando a SNR física do canal se degrada severamente. Isso ocorre pois o modelo foi treinado para ser resiliente a variações no espaço latente, garantindo que o *significado* seja transmitido de forma íntegra mesmo sob condições de canal que invalidariam transmissões baseadas puramente em bits.

F. Análise de Sensibilidade do Hiperparâmetro α O balanço entre as tarefas de classificação $\mathcal{L}_{CE}$ e de reconstrução $\mathcal{L}_{MSE}$, controlado por α , define a natureza do espaço latente. Observou-se que valores muito baixos de α resultam em espaços latentes excessivamente abstratos, o que dificulta a reconstrução visual, mesmo que a classificação seja mantida. Por outro lado, α elevado prioriza a fidelidade dos pixels, o que pode introduzir redundância indesejada no vetor latente. Nos experimentos conduzidos, o valor $\alpha = 0,5$ mostrou-se o

ponto ideal para garantir que a semântica seja útil tanto para máquinas quanto para humanos.

VII. LIMITAÇÕES E TRABALHOS FUTUROS

Apesar dos resultados promissores, este estudo abre portas para novas investigações. A limitação atual de 3 rodadas federadas fornece uma visão de convergência inicial; estudos de longo prazo são necessários para observar a estabilidade em cenários Non-IID severos. Além disso, a quantização dos vetores latentes de 32 bits para formatos de ponto fixo ou mesmo binários (Quantized VAEs) pode desbloquear ordens de magnitude adicionais de compressão. Futuras iterações incluirão a seleção adaptativa da dimensão latente em tempo real e o uso de modelos generativos mais profundos no receptor para reconstruções fotorrealistas a partir de sementes semânticas mínimas.

VIII. CONCLUSÃO

Este trabalho validou uma arquitetura de comunicação semântica federada capaz de operar com compressão extrema sem comprometer a eficácia da tarefa de classificação. Demonstra-se que autoencoders convolucionais leves são ferramentas poderosas para extração de significado localmente na origem dos dados e que o ruído de canal, tradicionalmente visto como um degradador de sinal, pode ser explorado estrategicamente como um mecanismo de regularização para criar representações latentes robustas. Nossos resultados experimentais no CIFAR-10 abrem caminho para sistemas 6G mais resilientes, sustentáveis e eficientes, em que a inteligência e o propósito da comunicação são priorizados em relação à fidelidade bruta dos bits transmitidos.

REFERÊNCIAS

- [1] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.
- [2] W. Yang, H. Du, Z. Liew, W. Y. B. Lim, Z. Xiong, D. Niyato, X. Chi, X. S. Shen, and C. Miao, "Semantic communications for future internet: Fundamentals, applications, and challenges," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 1, pp. 213–250, 2023.
- [3] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y. Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. International Conference on Artificial Intelligence and Statistics (AISTATS)*, pp. 1273–1282, 2017.
- [4] A. Krizhevsky, G. Hinton, *et al.*, "Learning multiple layers of features from tiny images," tech. rep., University of Toronto, 2009.
- [5] X. Luo, H.-H. Chen, and Q. Guo, "Semantic communications: Overview, open issues, and future research directions," *IEEE Wireless Communications*, vol. 29, no. 1, pp. 68–76, 2022.
- [6] H. Xie, Z. Qin, G. Y. Li, and B.-H. Juang, "Deep learning enabled semantic communication systems," *IEEE Transactions on Signal Processing*, vol. 69, pp. 2663–2675, 2021.
- [7] E. Grassucci, S. Barbarossa, and D. Communiello, "Generative AI meets semantic communication: Evolution, taxonomy, and prospects," *arXiv preprint arXiv:2310.08447*, 2023.
- [8] A. A. Alemi, I. Fischer, J. V. Dillon, and K. Murphy, "Deep variational information bottleneck," *CoRR*, vol. abs/1612.00410, 2016.