



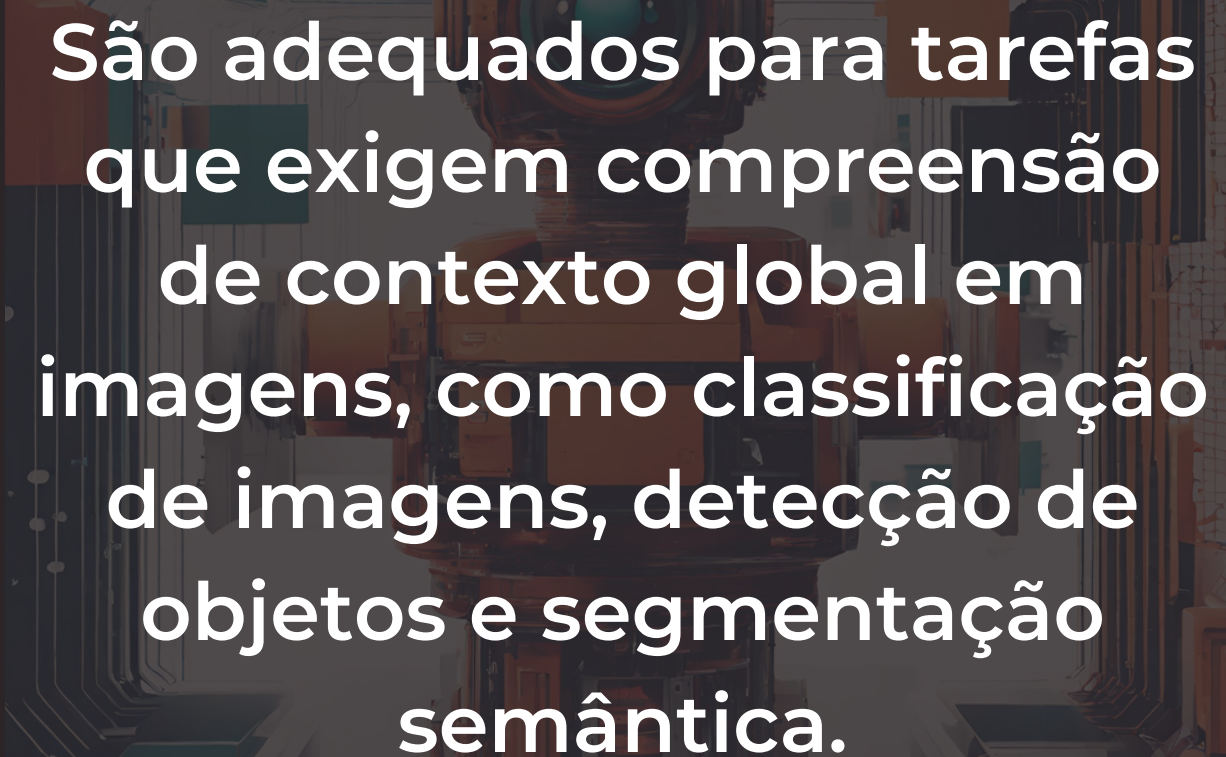
COMPUTER — VISION — SPRINT 4

95985 - Henrico Nardelli Bela
96232 - Felype Nunes de Souza
96302 - Sara Leal De Moura Lessa
94437 - Emily Gabrielly Alves Bispo dos Santos
94026 - Daniel Faria de Barros

Introdução

Visual Transformers, extensões da técnicas de Transformers, capazes de captar a atenção multi-cabeça para capturar relações entre pixels em uma imagem.

A imagem é dividida em patches, e esses patches são tratados como sequências de entrada para o modelo.



São adequados para tarefas que exigem compreensão de contexto global em imagens, como classificação de imagens, detecção de objetos e segmentação semântica.

Desenvolvimento

Casos de uso

- Classificação de Imagens
- Detecção de Objetos
- Segmentação Semântica
- Compreensão de Documentos Visuais

Para tarefas de detecção de objetos, como identificação da localização e classe de objetos em uma imagem.

Eles permitem que o modelo compreenda a cena global e as relações entre os objetos.



CLASSIFICAÇÃO

Modelos Transformers estão promissores para classificar objetos em imagens, incluindo a distinção entre Fanta Laranja e Coca-Cola.

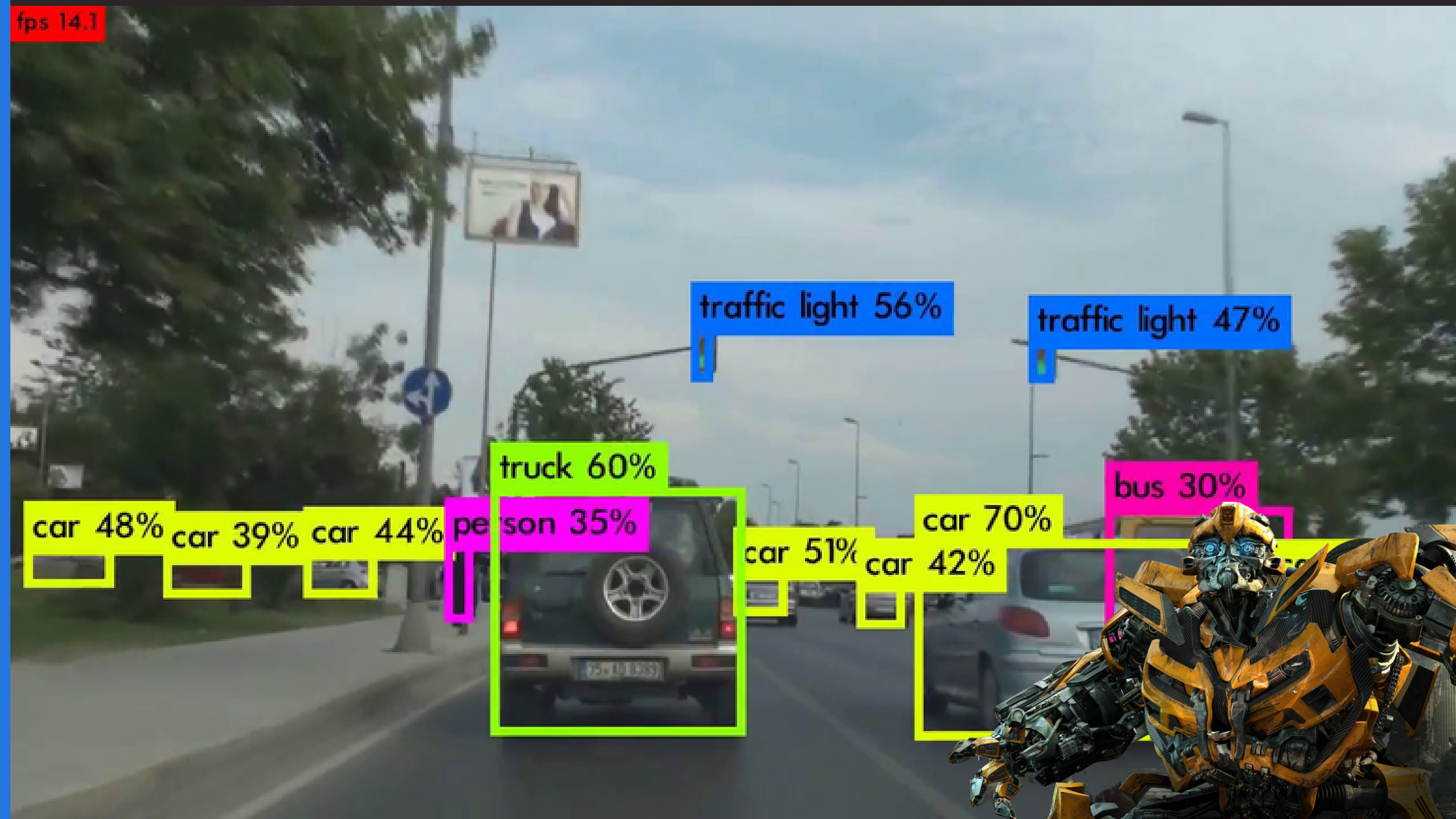
RELAÇÕES COMPLEXAS

A capacidade dos Transformers de entender relações complexas em dados sequenciais melhora significativamente a precisão e a eficiência no reconhecimento de objetos.

AVANÇOS FUTUROS

Espera-se que essas capacidades dos Transformers conduzam a avanços empolgantes na visão computacional e em outras áreas, transformando a maneira como lidamos com dados visuais e impulsionando a inovação em IA.

Destques



Reconhecimento de Objetos com Transformers:

A tarefa de reconhecimento de objetos em imagens envolve identificar e classificar objetos específicos dentro de uma cena. Tradicionalmente, isso era feito com abordagens baseadas em redes neurais convolucionais (CNNs). No entanto, os Transformers mostraram grande potencial neste domínio.

Aplicação na Tarefa de Reconhecimento de Fanta Laranja e Coca-Cola:

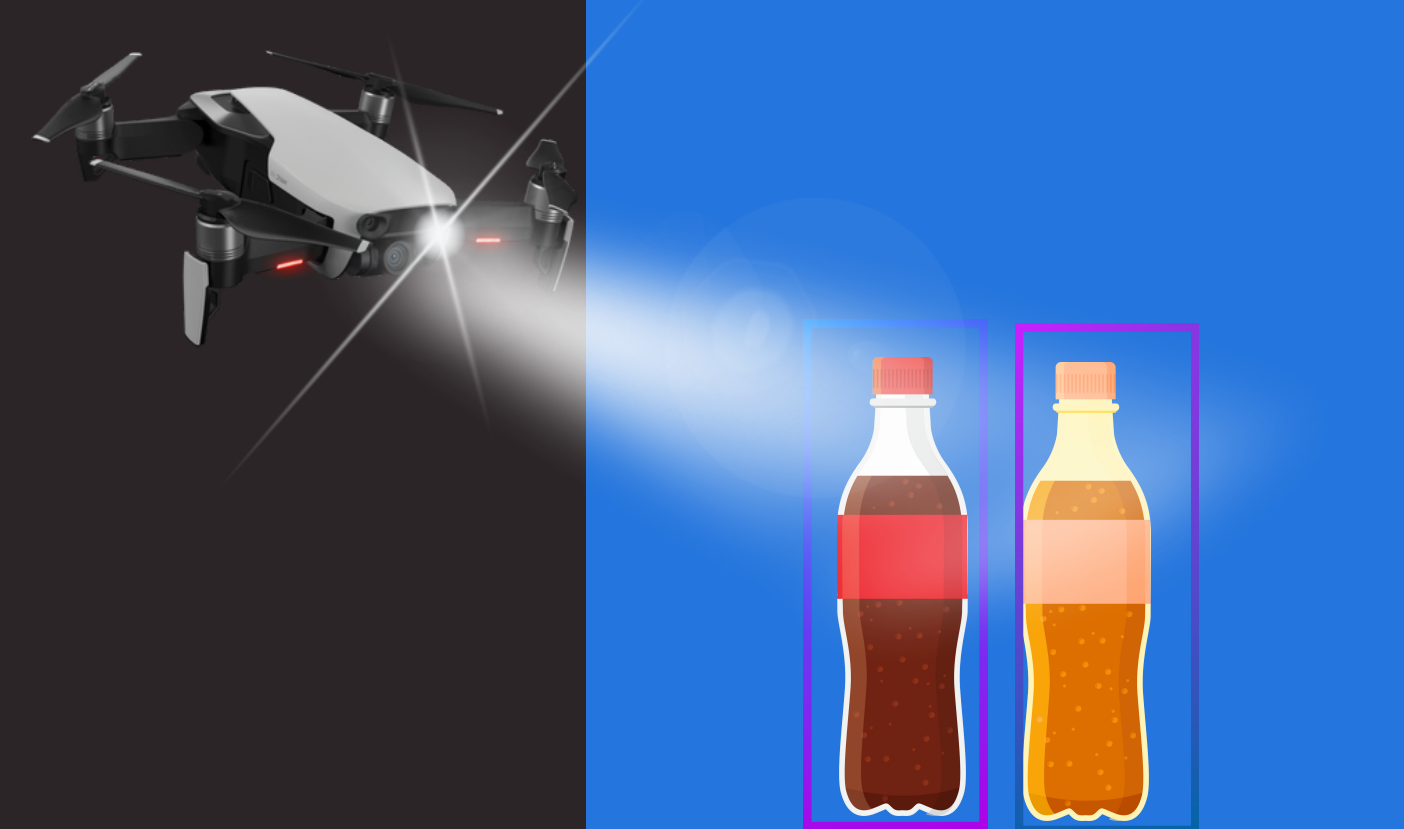
Agora, vamos considerar a aplicação prática dos Transformers na distinção entre Fanta Laranja e Coca-Cola a partir de imagens. Para fazer isso, precisamos seguir alguns passos:

1. Coleta de Dados:

Primeiro, é necessário criar um conjunto de dados contendo imagens rotuladas de Fanta Laranja e Coca-Cola. Quanto mais variadas forem as imagens, melhor.

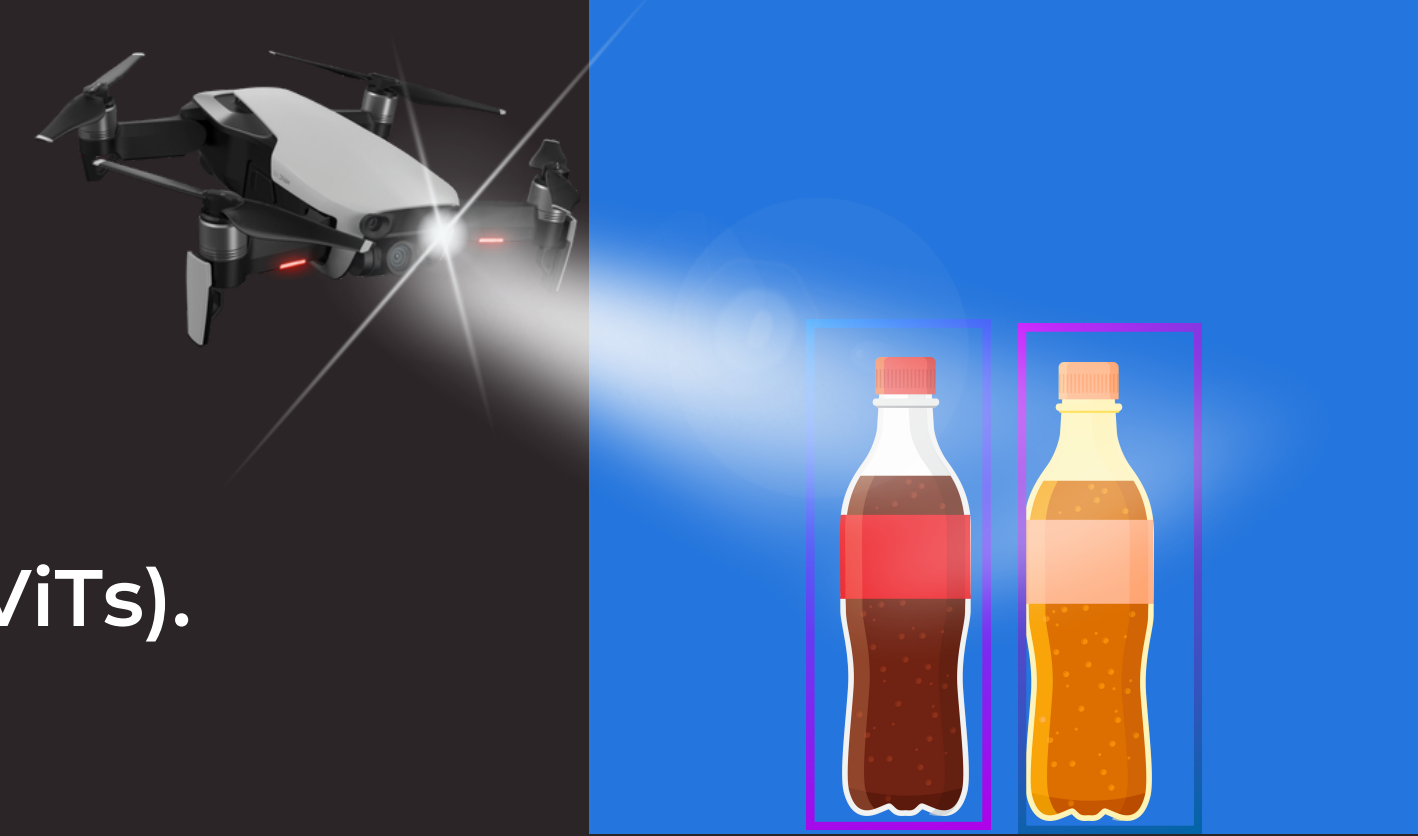
2. Pré-processamento de Dados:

As imagens devem ser pré-processadas para garantir que todas tenham o mesmo tamanho e formato, facilitando o processamento pelo modelo.



3. Escolha do Modelo Transformer:

Existem vários modelos Transformer pretreinados disponíveis, como BERT, GPT, e Vision Transformers (ViTs). Para tarefas de visão, os ViTs têm se destacado.



4. Fine-Tuning:

O modelo preexistente deve ser fine-tunado usando o conjunto de dados de Fanta Laranja e Coca-Cola. Isso permite que o modelo se ajuste à tarefa específica de reconhecimento.

5. Teste e Avaliação:

Agora, podemos testar o modelo em imagens não vistas para avaliar sua capacidade de distinguir as duas bebidas.

Conclusão

Os modelos Transformers têm uma aplicação promissora em tarefas de reconhecimento de objetos em imagens, como o reconhecimento de Fanta Laranja e Coca-Cola.

Sua capacidade de capturar relações complexas em dados sequenciais oferece um potencial significativo para melhorar a precisão e a eficiência nesse domínio.

À medida que continuamos a explorar as capacidades desses modelos, esperamos avanços emocionantes na visão computacional e em muitas outras áreas.

Referencias

<https://arxiv.org/abs/2206.06323>

<https://arxiv.org/abs/2101.11605>

<https://arxiv.org/abs/2203.12944>