Resumo: Microsoft Coco - Common objects in context

De Matheus Loiola Pinto Curado Silva

25 de setembro de 2023

Sumário

Introdução

Compreender cenas é um processo que possui inúmeras etapas, visto que é necessário compreender os atributos da cena, a relação entre eles, e o desenvolvimento de uma semântica para descrevê-los. Um exemplo que ajudou nessa compreensão é o *dataset* *ImageNet*, que possibilitou inúmeras descobertas nas áreas de classificação e detecção de objetos.

Assim, o artigo introduz um novo *dataset*, que é focado na **detecção visual de objetos, contextualização entre objetos e a localização precisa deles em imagens não icônicas**, para que seja possível avançar nas pesquisas de detecção de contexto e cenas, sem focar em apenas objetos isolados (imagens icônicas).

Dessa forma, para atingir esse objetivo, foram agrupadas imagens contendo relações contextuais e visualizações não icônicas, ou seja, os objetos estão fora do foco na imagem. O *dataset* possui 91 categorias com 82 delas contendo mais do que 5000 imagens. A ideia do *dataset* também é ter menos categorias, mas ter mais instâncias por categoria.

Trabalho relacionado

Os *datasets* relacionados com detecção de objetos podem ser divididos em três grupos: aqueles que se referem a classificação de objetos, detecção de objetos ou rotulagem semântica da cena.

- **Classificação de objetos**: A tarefa de classificação de objetos requer rótulos binários indicando se os objetos são presente em uma imagem. O *dataset* utilizado para isso pode ser o **ImageNet**, contendo mais de 22.000 categorias de objetos.

- **Detecção de objetos**: Detectar um objeto envolve tanto afirmar que um objeto pertencente a uma classe especificada e se ele está presente na imagem. O *dataset* utilizado para isso pode ser o **PASCAL VOC**.

- **Rotulagem semântica de cenas**: A tarefa de rotular objetos semânticos em uma cena requer que cada pixel de uma cena imagem seja rotulada como pertencente a uma categoria, como céu, cadeira, chão... O *dataset* utilizado para isso pode ser o **SUN**.

Coleção de imagens

O *dataset* tem o foco em objetos, ou seja, pessoas, carros, cadeiras, etc. e não possui foco em coisas que geralmente não tem limites, como ruas, grama ou céu. Além disso, as categorias são limitadas a termos básicos, frequentemente usados por humanos, como utilizar a categoria "cachorro"ao invés de "pastor-alemão"ou outras raças.

Anotações de imagens

Essa seção demonstra o processo para criar o *dataset MS-COCO*.

A primeira etapa foi definir quais categorias estavam presentes em cada imagem. Após isso, todas as instâncias dos objetos de certa categoria são rotuladas. Dessa forma, o estagio final é segmentar cada instância do objeto na imagem. Apos essas três etapas, os pesquisadores analisam os resultados e verificaram quais imagens precisariam de revisão. Por fim, foram adicionadas cinco legendas para cada imagem no *dataset*.

Estatísticas e resultados finais

O *MS COCO* foi desenvolvido para detecção e segmentação de objetos em seu contexto natural. O *dataset* foi separado em 50% das imagens para treinamento, 25% para validação e teste.

É notável, também, que em comparação com o *PASCAL VOC*, existe uma diferença de performance entre o *dataset MS COCO* ao treinar uma rede neural, visto que este último possui imagens mais difíceis (não icônicas) nas quais os objetos estão parcialmente ocultos.