Achei que a contextualização poderia ter sido melhor explorada.

Contextualização foi expandida.

Quais foram os tratamentos (pré-processamento) realizados na sua base de dados? O aluno basicamente só falou da divisão da base.

Subdividi a seção “Coleta de Dados” em “Estrutura da base” e “Pré-processamento”, remanejando e ampliando parte do texto que estava em outra seção.

Se existiam 6171 patches (3.621 vagas ocupadas e 2.550 vagas vazias), por que uso apenas 100 para o treinamento? Muito pouco.

Treinamento aqui talvez não seja o termo correto, pois é usada uma rede pré-treinada, a ResNet18. Essas 100 imagens são usadas como base, ou exemplo, de vagas vazias e ocupadas. Cada imagem de busca será comparada contra cada uma dessas 100 e a mais semelhante será a que definirá o estado da vaga na imagem de busca (vazia/ocupada). Portanto quanto maior a quantidade dessas imagens de exemplo maior o tempo de processamento. Com 100 imagens o tempo para cada imagem de busca ficou em 2,3 segundos, o que já é bastante alto. Com menos imagens eu teria perda na acurácia. Portanto 100 me pareceu um valor equilibrado. Troquei o termo “dataset de treinamento” por “dataset de exemplo”, para evitar esse mal entendido e fiz adendos no texto definindo os termos e adicionando esta explicação.

Em Análise e Exploração de Dados esperava "visualizar" a base de dados (lembre-se da disciplina Visualização de Dados). Não entendi muito bem o que o aluno quis fazer pois em algumas partes desta seção ele fala de atividades de pré-processamento.

O subtítulo neste caso está errado, não se trata de pré-processamento de imagens e sim uma explicação geral sobre a CNN ResNet18, que inclui o pré-treinamento dela na base de imagens ImageNet. Mudei o subtítulo para “Visão geral sobre a ResNet18”.

O aluno afirma que a resnet18 possui parâmetros fixos, por ser uma rede pré-treinada. Estes parâmetros deveriam ter sido apresentados no texto.

[1] Existia uma parte do texto sobre a arquitetura da resnet18 e seus parâmetros (tamanho das imagens, normalização). Expandi esta parte, incluindo mais informações sobre as camadas da CNN e uma nova referência (JAY, 2018).

O aluno deveria ter apresentado "recortes" do código e feito as devidas explicações no texto, não apenas colocado o link para o código completo.

Fiz a apresentação de trechos de cada passo do script, com texto explicativo, está na seção 3 “Processamento/Tratamento de Dados”.

Em momento algum ele apresenta a arquitera da rede convolucional utilizada. Não é possível compreender como ela está estruturada.

Mesmo caso do item marcado com [1].

Deveria ter apresentado a fórmula do cálculo de similaridade, não apenas relatar que usou o cosseno.

Expandi a referida seção, usando diagrama e fórmulas vistos em (KORZEC, 2020) com as devidas explicações. Também adicionei uma referencia para um livro de Sistemas de Recomendações que pode ser interessante ao leitor para aprofundamento (AGGARWAL, 2016).

Na página 5 o aluno fala de acurácia. Não sei se ele pretendia usar esta métrica no trabalho, pois não encontrei no texto.

Acurácia seria o índice de acertos das identificações de vagas vazias ou ocupadas. É expressado em forma de percentual na seção Apresentação de Resultados. Fiz um adendo ao texto para esclarecer essa questão na subseção 1.2 (página 5).

Na página 13 ele fala de 6071 imagens, mas não eram 6171?

Na verdade, as 6.071 imagens são as que sobraram após as 100 imagens de treinamento (agora renomeadas para imagens de exemplo) terem sido retiradas do conjunto original de 6..171. Fiz um adendo ao texto colocando essa observação (página 7, subseção 2.2).

O que é um "dataset de busca", como citado na página 13?

O dataset de busca seria as 6071 imagens que simulam o caso de uso real, ou seja, cada uma delas é comparada contra cada uma das 100 imagens do “dataset de exemplo” para a definição do estado (vazia/ocupada). Fiz um adendo ao texto clarificando essas definições (página 7, subseção 2.2).

O aluno não apresenta métricas de erro para que possamos avaliar o resultado da rede.

Existe a métrica de acurácia (taxa de acerto) em percentual calculada para ambos os datasets. Incluí uma tabela ao final da seção “Apresentação de resultados” consolidando os dados.

Deveria ter realizado testes com outros modelos de redes convolucionais, apresentar os resultados da métricas em tabelas como forma de comparar o desempenho de cada um.

Repeti o experimento com as outras redes da “família” ResNet: 34, 50, 101 e 152. O resultado coloquei em uma nova subseção no final da seção “Apresentação de Resultados”.

Achei a apresentação dos resultados muito básica.

Incluí duas tabelas e um gráfico na seção, para demonstração do desempenho da ResNet18 e outras redes convolucionais, em termos de acurácia (taxa de acerto) e performance (tempo de execução).

Também expandi a seção “Usando listas de top-k mais semelhantes” discorrendo sobre a escolha do valor k=4 e a utilidades das k imagens mais semelhantes, ao invés de usar apenas uma.

Fábio Martins de Oliveira

fmartinssp@yahoo.com.br