PSI5790 Aprendizado Profundo para Visão Computacional
Primeiro período de 2025 Exercício-programa
Data de entrega: 03/06/2025 (terça-feira) até 23:59 horas
Nota: Cada dia de atraso acarreta um desconto de 1 ponto na nota. Não pode entregar com mais de 7 dias de atraso.

O livro [*Dive into Deep Learning*], disponível gratuitamente online, fornece um conjunto de imagens denominado *banana-detection*, útil para experimentos com conceitos de detecção de objetos.
https://d2l.ai/chapter_computer-vision/object-detection-dataset.html
Esse conjunto é composto por 1.000 imagens de treino e 100 de teste, todas coloridas e com resolução de 256×256 pixels. Em cada imagem, uma banana foi inserida em posições, tamanhos e orientações aleatórias (Figs. 1a e 1c).

Para enriquecer o experimento, foram adicionadas mais 1.000 imagens de treino e 100 de teste provenientes do conjunto Pascal VOC, que não contêm bananas.
http://host.robots.ox.ac.uk/pascal/VOC/
A seleção consistiu em recortar a região central quadrada de cada imagem e redimensioná-la para 256×256 pixels, a fim de manter a consistência com o conjunto original (Figs. 1b e 1d).


(a) Imagem de treino 0.png,1,104,20,143,58


(b) Imagem de treino 70032.jpg,0,0,0,0,0


(c) Imagem de teste 0.png,1,183,63,241,112


(d) Imagem de teste 90001.jpg,0,0,0,0,0

Figura 1: As imagens (a) e (b) correspondem ao conjunto de treino, enquanto (c) e (d) pertencem ao conjunto de teste. Uma banana foi inserida nas imagens (a) e (c); as imagens (b) e (d) não contêm bananas.

Dois arquivos denominados *label.csv* (um para o conjunto de treino e outro para o de teste) contêm as coordenadas da *bounding box* da banana em cada imagem. A estrutura do arquivo segue o formato:

```
img_name,label,xmin,ymin,xmax,ymax
0.png,1,104,20,143,58
1.png,1,68,175,118,223
2.png,1,163,173,218,239
3.png,1,48,157,84,201
(...)
70027.jpg,0,0,0,0,0
70032.jpg,0,0,0,0,0
70033.jpg,0,0,0,0,0
70039.jpg,0,0,0,0,0
(...)
```

Cada linha representa uma imagem, iniciando com o nome do arquivo, seguido pelo rótulo da classe (label) e pelas coordenadas da bounding box no formato (*xmin*, *ymin*, *xmax*, *ymax*). O rótulo 1 indica a presença do objeto de interesse (neste caso, uma banana), enquanto 0 indica a ausência de qualquer objeto de interesse. Para imagens sem objeto, as coordenadas são registradas como (0, 0, 0, 0) por convenção, embora possam ser ignoradas na prática.

Para baixar este conjunto do Google Drive para computador local, primeiro instale/atualize *gdown*:

```
$ pip3 install --upgrade gdown
```

Nota: Não é necessário fazer isto no Google Colab, pois *gdown* está pré-instalada.

Depois, execute o código abaixo (tanto no Colab como no computador local), que copiará o arquivo *BanPas.zip* do Google Drive para o seu diretório atual e o descompactará.

```
import os; import gdown
nomeArq="BanPas.zip"
if not os.path.exists(nomeArq):
  os.system("gdown 17GzJnCrIQNXO5qwq7pnI_y0Fa-VlHbN4")
os.system("unzip -u "+nomeArq)
```

O objetivo deste exercício é implementar um único modelo de rede neural capaz de resolver dois problemas ao mesmo tempo:

a) Dizer se tem (ou não) uma banana na imagem.

b) Se houver banana, achar a sua localização (*xmin*, *ymin*, *xmax*, *ymax*).

A rede neural deve receber uma imagem e gerar 5 saídas: presença/ausência de banana, e as 4 coordenadas da banana.

Importante: Você obrigatoriamente deve usar uma única rede neural para resolver este problema. Não pode usar uma rede para detectar se tem ou não banana e outra para localizar banana na imagem.

Evidentemente, você **não** pode usar um modelo pronto de **detecção** de objetos. Porém, pode usar todas as outras técnicas que aprendemos no curso. Inclusive, pode usar um modelo de **classificação** de imagens pré-treinado em ImageNet (como VGG, ResNet, EfficientNet, etc.) e usá-lo como modelo-base.

Informe no relatório e no vídeo:

1) A taxa de erro em classificar imagem como com/sem banana.

Nota: A minha implementação cometeu 1 único erro (0,5% de erro) ao dizer que imagem 44.png não tem banana quando na verdade tem.	 Figura 2: 44.png com banana.

2) Nas imagens com banana que o seu programa identificou corretamente como contendo banana, o erro médio absoluto em pixels entre as coordenadas verdadeiras e as coordenadas preditas pelo seu programa.

Nota: O erro médio absoluto do meu programa foi 2,65 pixels. Isto é, o meu programa reconheceu banana em 99 das 100 imagens com banana. Assim, foi calculada a diferença absoluta média entre 4×99 coordenadas verdadeiras e 4×99 coordenadas preditas pelo programa.

A nota do EP dependerá desses dois erros. A figura 3 mostra algumas saídas do meu programa.

Coloque no relatório uma imagem semelhante à figura 2 mostrando as detecções das bananas pelo seu programa nas 12 primeiras imagens de teste com/sem banana.

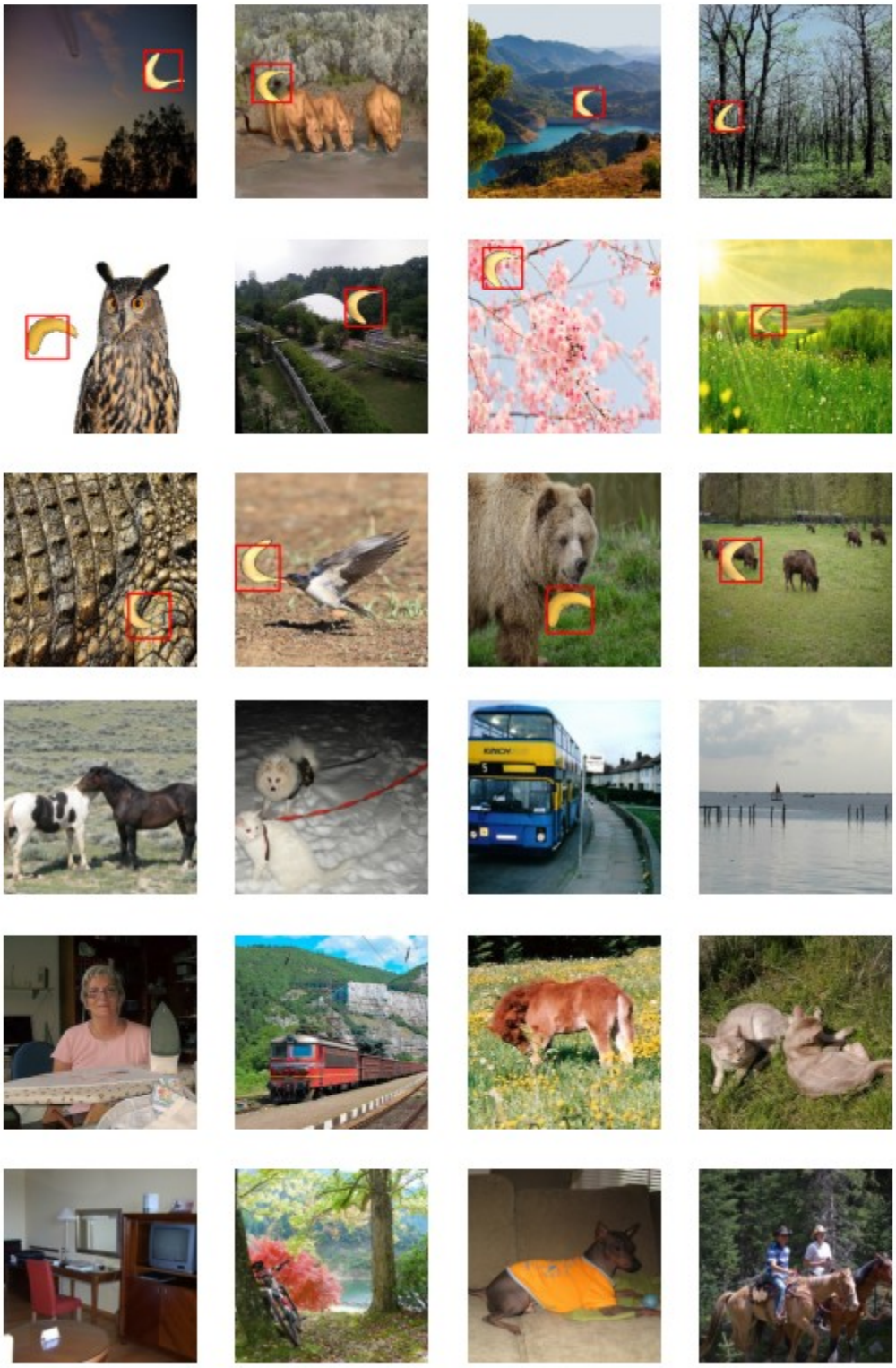

Figura 3: Saídas geradas pelo meu programa nas 12 primeiras imagens de teste com e sem banana.

Obs. 1: Este exercício deve ser resolvido individualmente.

Obs. 2: Em princípio, deve usar o mesmo ambiente de programação utilizado em aula (Python, Keras, OpenCV), computador local ou Google Colab. Se quiser utilizar outro ambiente, converse antes com o professor. Se alguém quiser programar em PyTorch, pode.

Obs. 3: Entregue os programas-fontes (banpas.py ou banpas.ipynb). Você pode enviar link para Colab em vez dos programas: neste caso, assegure que o professor (hae.kim@usp.br) tenha acesso. Não entregue o dataset.

Obs. 4: Entregue um documento PDF de no máximo 4 páginas (relatorio.pdf) descrevendo o funcionamento do seu programa, os resultados obtidos e as suas conclusões. O envio do relatório é obrigatório (veja o anexo). O relatório é um documento Word/LaTex/LibreOffice convertido para PDF. Um notebook Python anotado não será aceito como relatório.

Obs. 5: Entregue um vídeo de no máximo 120s explicando o funcionamento do seu programa, os resultados obtidos e as suas conclusões. No vídeo deve aparecer em algum momento o seu rosto e um documento seu. É necessário que o vídeo contenha áudio, pois é muito difícil entender um vídeo sem a explicação falada. O vídeo não pode ter sido acelerado para diminuir a duração, pois dificulta entender a fala. Pode enviar o vídeo em si (.mkv, .mp4, .avi, etc.) ou um link para o vídeo (youtube, google drive, etc). No segundo caso, assegure que o professor tenha acesso.

Obs. 6: Envie o material através de edisciplinas. Dentro do prazo, você pode substituir o material anterior por um novo.

Anexo: Relatórios dos exercícios programas

O mais importante numa comunicação escrita é que o leitor entenda, sem esforço e inequivocamente, o que o escritor quis dizer. O texto ficar “bonito” é um aspecto secundário. Se uma (pseudo) regra de escrita dificultar o entendimento do leitor, essa regra está indo contra a finalidade primária da comunicação. No site do governo americano [1], há regras denominadas de “plain language” para que comunicações governamentais sejam escritas de forma clara. As ideias por trás dessas regras podem ser usadas em outros domínios, como na escrita científica. Resumo abaixo algumas dessas ideias.

(1) Escreva para a sua audiência. No caso do relatório, a sua audiência será o professor ou o monitor que irá corrigir o seu exercício. Você deve enfocar na informação que o seu leitor quer conhecer. Não precisa escrever informações que são inúteis ou óbvias para o seu leitor.

(2) Organize a informação. Você é livre para organizar o relatório como achar melhor, porém sempre procurando facilitar o entendimento do leitor. Seja breve. Quebre o texto em seções com títulos claros. Use sentenças curtas. Elimine as frases e palavras que podem ser retiradas sem prejudicar o entendimento. Use sentenças em ordem direta (sujeito-verbo-predicado).

(3) Use o tempo verbal o mais simples possível. Evite “verbos ocultos” (por exemplo, substitua “precisamos realizar uma revisão das contas” para “precisamos rever as contas”; substitua “fiz o pagamento do seu salário” por “paguei o seu salário”). Evite cadeia longa de nomes, substituindo-os por verbos (em vez de “desenvolvimento de procedimento de proteção de segurança de trabalhadores de minas subterrâneas” escreva “desenvolvendo procedimentos para proteger a segurança dos trabalhadores em minas subterrâneas”). Minimize o uso de abreviações (para que o leitor não tenha que decorá-las). Use sempre o mesmo termo para se referir à mesma realidade, pois pode confundir o leitor se usar termos diferentes para se referir a uma mesma coisa. O relatório não é obra literária, não tem problema repetir várias vezes a mesma palavra.

(4) Use voz ativa. Deixe claro quem fez o quê. Se você utilizar oração com sujeito indeterminado ou na voz passiva, o leitor pode não entender quem foi o responsável (Ex: “Criou-se um novo algoritmo” - Quem criou? Você? Ou algum autor da literatura científica?). O site diz: “Passive voice obscures who is responsible for what and is one of the biggest problems with government writing.”

(5) Use exemplos, diagramas, tabelas, figuras e listas. Ajudam bastante o entendimento.

1 https://plainlanguage.gov/guidelines/