📘Image Captioning and Segmentation Project Plan (Full Pipeline)

📁Project Directory: `/Volumes/ExternalHDD/ICP/`

🔖Folder Structure

```
ICP/
├── data/
│   ├── MSCOCO/
│   │   ├── annotations/
│   │   ├── train2017/
│   │   └── val2017/
│   └── PascalVOC/ (optional)
│
├── notebooks/
│   ├── 01_EDA.ipynb
│   ├── 02_Captioning_Model.ipynb
│   ├── 03_Segmentation_Model.ipynb
│   ├── 04_Integration.ipynb
│   └── 05_Evaluation.ipynb
│
├── src/
│   ├── captioning/
│   │   ├── model.py
│   │   ├── dataset.py
│   │   └── train.py
│   ├── segmentation/
│   │   ├── unet.py / mask_rcnn.py
│   │   ├── dataset.py
│   │   └── train.py
│   └── utils/
│       ├── preprocessing.py
│       ├── evaluation.py
│       └── visualization.py
│
├── checkpoints/
│   ├── captioning_model/
│   └── segmentation_model/
│
├── app/
│   ├── app.py
│   └── templates/
│
```

```
├── README.md
└── requirements.txt
```

Project Timeline & Steps

Week 1–2: Research

- Study image captioning (CNN+LSTM, Transformers)
- Study segmentation (U-Net, Mask R-CNN)
- Read: CS231n, Karpathy blog, papers (Show and Tell, U-Net, Mask R-CNN)

Week 3–4: Setup

- Create virtual environment: `virtualenv venv`
- Install: `tensorflow`, `torch`, `opencv-python`, `nltk`, `streamlit`, etc.
- Download MS COCO data into `data/MSCOCO/`

Week 5–6: Image Captioning Model

- Implement CNN encoder (custom ResNet)
- Build LSTM decoder from scratch
- Prepare vocabulary using NLTK
- Train model and save to `checkpoints/captioning_model/`

Week 7–8: Image Segmentation

- Implement U-Net or Mask R-CNN from scratch
- Train with MS COCO masks
- Save model to `checkpoints/segmentation_model/`

Week 9: Integration

- In a notebook, run captioning + segmentation on same image
- Visualize side-by-side outputs

Week 10–12: Deployment

- Evaluate BLEU, IoU, Dice scores
- Build Streamlit app under `app/`
- App workflow: Upload image → Generate caption → Show segmented image

---

🔗 Deliverables

- Full GitHub repo
- Jupyter notebooks and Python scripts
- Final trained weights
- Web demo (optional)
- Report + demo video

🔍MacBook M2 (8GB) Tips

- Use small batch sizes
- Train models in stages
- Optimize dataloader pipelines
- Mixed-precision or CPU-only training

🔗What's Next?

Let me know what you need: - Code templates (CNN+LSTM, U-Net)? - Dataset preprocessing scripts? - README.md and report structure? - Streamlit app base code?

Ready when you are! 💻