

Sample Project Ideas

1. Monocular Depth Estimation (Predict dense depth maps from a single RGB image)

- Dataset: NYU Depth V2 (Kaggle version) or KITTI Depth Dataset (Kaggle)
- Implement MiDaS (DPT) or DenseDepth with lightweight transformers. Compare CNN (ResNet encoder) vs Vision Transformer encoders. Use pretrained MiDaS weights and fine-tune on limited scenes.

2. Image Inpainting / Masked Image Reconstruction (Reconstruct masked or occluded image regions)

- Dataset: CelebA Faces (Kaggle)
- Implement MAE (Masked Autoencoder) or Partial Convolution U-Net. Compare CNN vs ViT-based vs pre-trained Stable Diffusion inpainting models.

3. Text Summarization with Transformers (Generate concise summaries from news articles)

- Dataset: BBC News Summary Dataset (Kaggle)
- Fine-tune T5-small or BART-base for abstractive summarization. Compare extractive vs abstractive methods.
- Evaluate with ROUGE and BERTScore. Explore attention heatmaps to analyze summary faithfulness.

4. Reduce deep CNN feature representations for visualization or classification.

- Extract CNN features then apply PCA, UMAP, or autoencoders. Train a classifier on reduced space. Compare Autoencoder vs Transformer-based reduction. Analyze feature interpretability in reduced spaces.

5. Scene Understanding with Lightweight Transformer

- Semantic segmentation + object counting.
- Use a lightweight transformer (SegFormer-Tiny) to segment urban scenes and estimate vehicle count.
- Dataset: Cityscapes small subset (Kaggle)
- Train or fine-tune SegFormer-Tiny or MobileNet DeepLabv3+. Postprocess masks to count instances.

6. Summarization & Sentiment Fusion (Mini NLP Project)

- Text Summarization + Sentiment Analysis. Build a model that summarizes news articles and classifies their sentiment or tone.
- Use T5-small or similar model for summarization.
- Use DistilBERT for sentiment classification.
- Combine pipeline outputs → “Summary + Mood” report.
- Dataset: BBC News Summary (Kaggle)

7. Pet Segmentation (Trimap-based / Matting)

- Dataset: Oxford-IIIT Pet Dataset (Kaggle)
- Segment and possibly extract alpha mattes for pet images.
- Train U-Net or Mask R-CNN for segmentation.
- Fine-tune MODNet or FBA Matting for hair/fur matting.

8. Efficient Real-Time Super-Resolution using a Residual Network.

- Implement a small, highly optimized convolutional neural network (e.g., a shallow ResNet-based architecture) for single-image super-resolution, also implement a Transformer based architecture focusing on achieving a good Peak Signal-to-Noise Ratio (PSNR) at very fast inference speeds. Compare both Resnet based and transformer-based models.

9. Road Object Detection for Autonomous Driving

- Dataset: BDD100K Mini (Kaggle) or a filtered COCO subset
- Goal: Detect vehicles, pedestrians, and traffic lights.
- Baseline: Train or fine-tune YOLOv5n / YOLOv8n or SSD-MobileNet (low compute). Add depth estimation (MiDaS) → “3D-aware bounding boxes.”