Car Model Classification

Computer Vision Slide – Group 15

Nguyen Truong Son, Tran Viet Cuong, Doan Duc Hoang,
Tran Trong Nghia, Nguyen Dam Quan

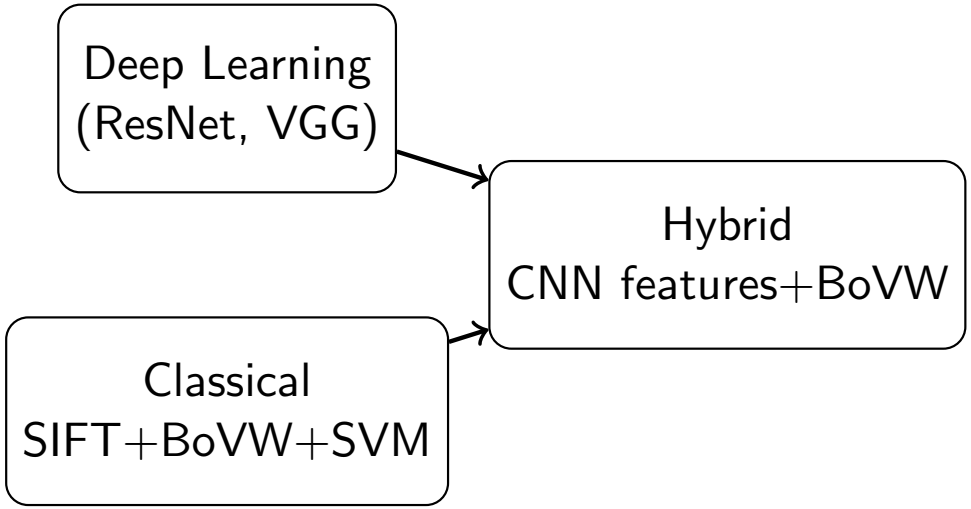
University of Science and Technology of Hanoi
Vietnamese France University
ICT Lab

May 2025

Problem Definition

- **Input**: An image of the car
- **Output**: Car make & model (8 classes)

Global & Existing Approaches

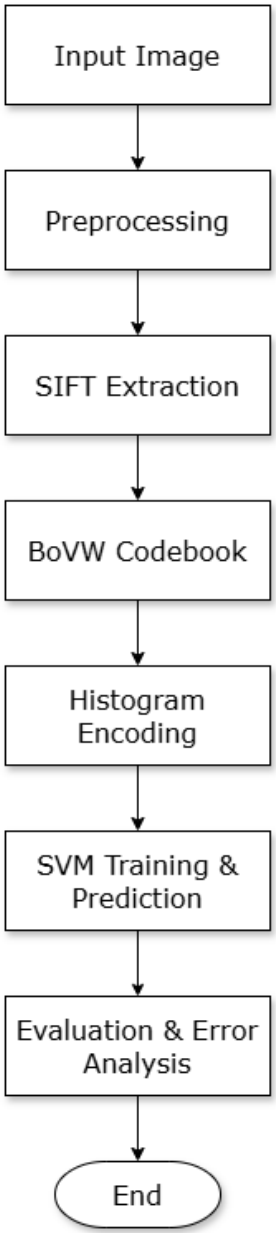


Common solutions:

- **Deep Learning**: fine-tuned CNNs (ResNet, VGG) — high accuracy, heavy compute
- **Classical**: SIFT + BoVW + SVM — interpretable, low resources
- **Hybrid**: CNN features + BoVW + SVM — balanced performance

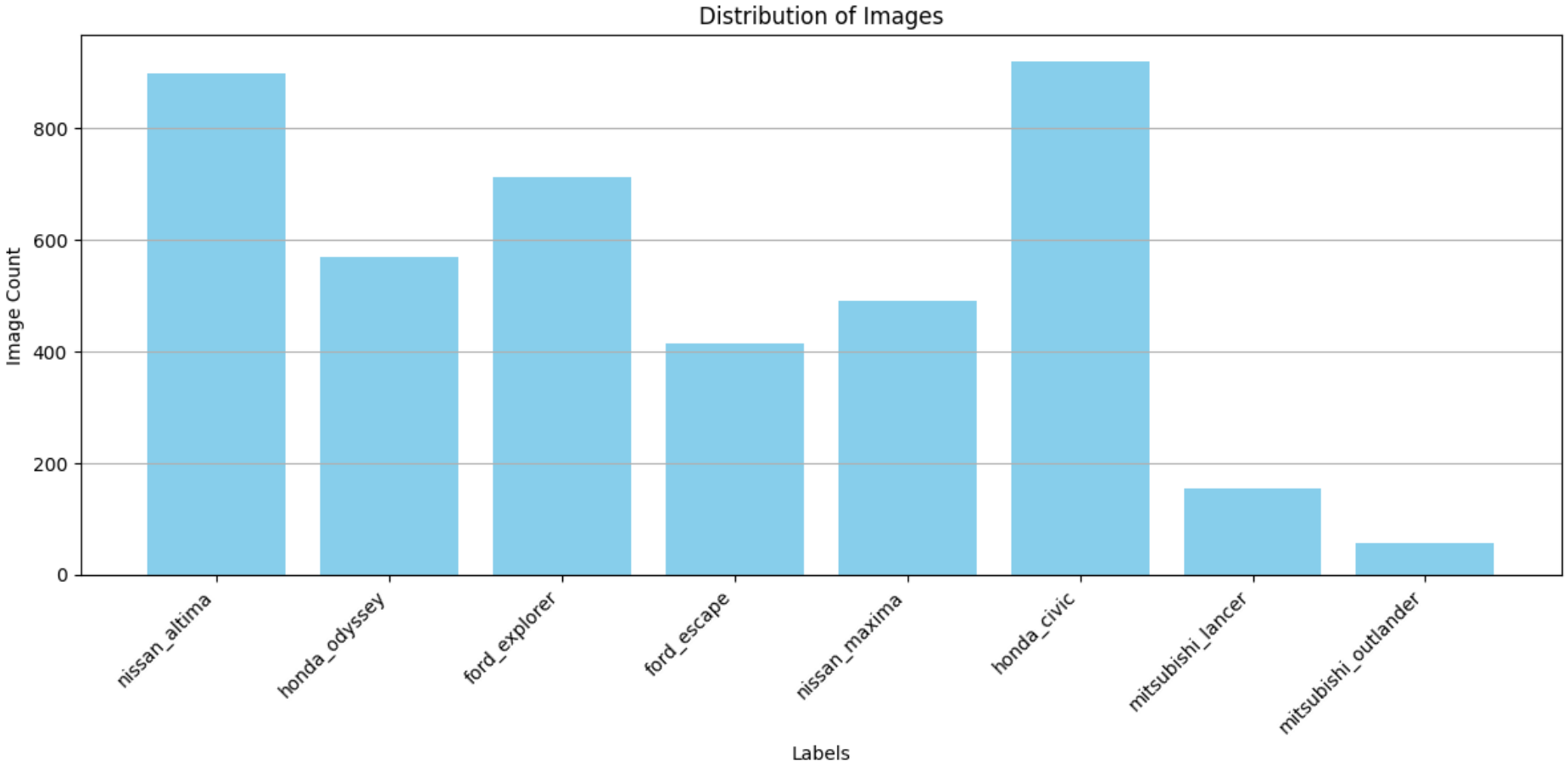
Summary from report Introduction and Related Work

Our Approach Overview



Data & Splitting

- Dataset: 4,216 images, 8 classes
- Method: 80% train/10% val/10% test

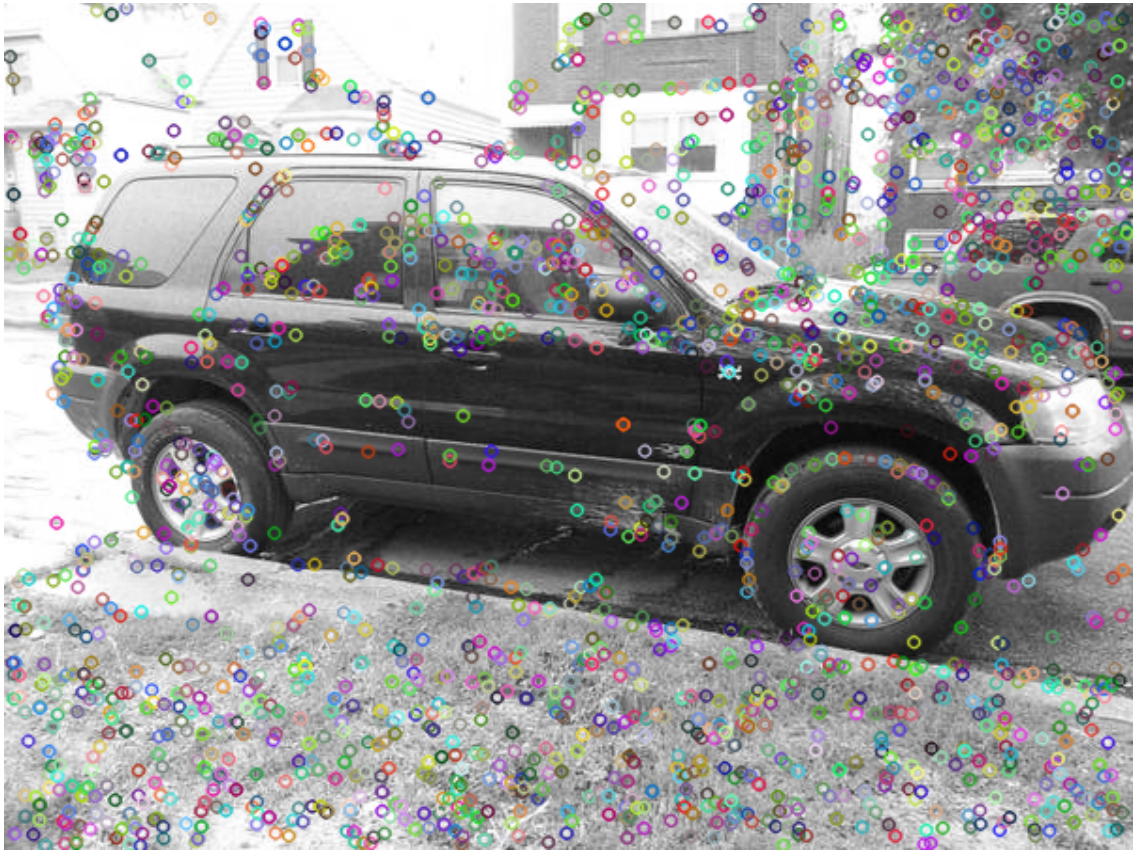


Preprocessing

- Resize to 256×256
- Convert to grayscale
- Label encoding

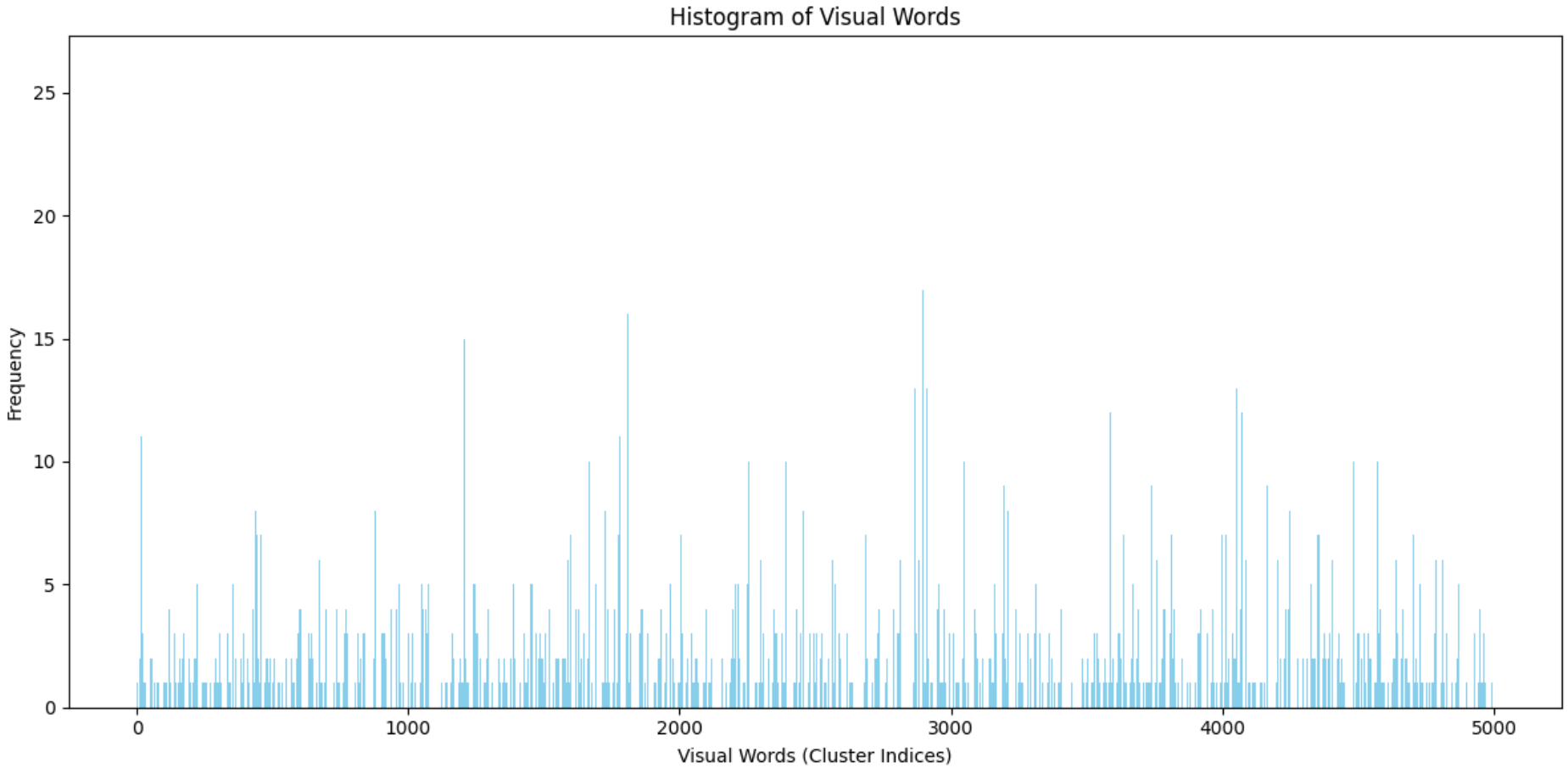
Feature Extraction (SIFT)

- Detect keypoints
- Compute 128-D descriptors



Codebook (BoVW)

- k-means clustering (k=5000 visual words)
- Map descriptors to histogram



Classification (SVM)

- Kernels: Linear, RBF, Sigmoid
- Training: 3 370 images
- Validation: 418 images
- Test: 428 images

Evaluation & Comparison

We evaluated our models on four key metrics and highlight the best results:

- **Metrics**: Accuracy, Precision, Recall, F1
- **Sigmoid SVM**: Accuracy 60%, F1 0.58
- **ResNet-18 fine-tune**: Accuracy 75%, F1 0.72

Error Analysis

Class	Precision	Recall
BMW	0.62	0.55
Audi	0.58	0.50
Toyota	0.65	0.60
Ford	0.60	0.57

Table: Per-class precision and recall from test results

Overlap between BMW and Audi features leads to misclassifications.

Conclusion & Future Work

- Classic pipeline plateau at 60% accuracy
- Kernel switching alone is not enough
- Current model can't be deployed for real-life applications
- Next steps:
  - Collect more data on imbalanced classes
  - Data augmentation & hyperparameter tuning
  - Ensemble methods

Questions?

The End