

**Proposed Master's Thesis**

**Improving a Compact Vision Model  
Using Knowledge Distillation**

**Student:** Gheith Alrawahi

**Student ID:** 2120246006

**Program:** Software Engineering

**Institution:** Nankai University

**Supervisor:** Prof. Jing Wang

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Project Objectives</b>	<b>2</b>
<b>3</b>	<b>Methodology and Model Specifications</b>	<b>2</b>
3.1	Model Selection . . . . .	2
3.2	Knowledge Distillation Setup . . . . .	3
3.3	Evaluation Dataset and Metrics . . . . .	3
<b>4</b>	<b>Expected Outcomes</b>	<b>3</b>
<b>5</b>	<b>References</b>	<b>4</b>

# 1 Introduction

Deep learning models in computer vision achieve impressive accuracy but often require large memory and computation resources, limiting deployment in real-time or edge environments. **Knowledge Distillation (KD)** provides an effective way to compress large models by transferring knowledge from a high-capacity **Teacher Model** to a smaller **Student Model**.

In this project, KD will be used to train a compact image classification model that achieves a balance between **accuracy and efficiency**. The objective is to improve the student model's performance while significantly reducing computational costs.

# 2 Project Objectives

The project will focus on the following achievable objectives:

1. **Baseline Establishment:** Train and evaluate both the Teacher (EfficientNetV2-L) and Student (EfficientNetV2-S) models using standard training to define baseline accuracy and resource requirements.
2. **Knowledge Distillation Implementation:** Apply standard KD techniques (soft targets + hard labels) to effectively transfer the teacher's knowledge to the student.
3. **Performance Analysis:** Quantitatively measure improvement in **accuracy**, **inference speed**, and **model size** between baseline training and KD training.
4. **Result Evaluation:** Analyze the effect of KD temperature, loss weighting, and dataset complexity on the student's learning efficiency.

# 3 Methodology and Model Specifications

## 3.1 Model Selection

Role	Model	Parameters (Approx.)	ImageNet Top-1 Accuracy (Approx.)
Teacher Model	EfficientNetV2-L	120 Million	85.7%
Student Model	EfficientNetV2-S	22 Million	83.9%

Table 1: Teacher and Student Model Specifications

**Framework:** PyTorch will be the main implementation platform.

### 3.2 Knowledge Distillation Setup

The total loss function will combine both teacher supervision and ground-truth learning:

$$L_{total} = \lambda_{KD} \cdot L_{KD} + \lambda_{CE} \cdot L_{hard} \quad (1)$$

Where:

- $L_{KD}$ : Kullback-Leibler divergence between the teacher's soft targets and the student's logits (temperature  $T = 4$ ).
- $L_{CE}$ : Standard cross-entropy loss with hard labels.
- The hyperparameters  $\lambda_{KD}$  and  $\lambda_{CE}$  will be optimized to balance soft and hard learning.

### 3.3 Evaluation Dataset and Metrics

**Dataset:** CIFAR-100 is selected for its manageable size and established benchmarking value. If time and resources allow, a subset of ImageNet-1K will be used for extended evaluation.

**Metrics:**

- **Top-1 Accuracy**
- **Validation Loss**
- **Inference Time**
- **Model Size**
- **Training Speed** (optional)

## 4 Expected Outcomes

The project is expected to deliver the following outcomes:

- The student model trained through KD will achieve **higher accuracy** than its independently trained baseline.
- The KD approach will demonstrate **better efficiency** in inference time and resource utilization compared to the teacher model.
- The final model will provide a strong **accuracy–efficiency trade-off**, proving KD as a practical compression strategy for vision models.

## 5 References

1. Hinton, G., Vinyals, O., & Dean, J. (2015). *Distilling the Knowledge in a Neural Network*. arXiv:1503.02531. <https://arxiv.org/abs/1503.02531>
2. Tan, M., & Le, Q. (2021). *EfficientNetV2: Smaller Models and Faster Training*. arXiv:2104.00298. <https://arxiv.org/abs/2104.00298>
3. Liu, Z., Mao, H., Wu, C., Feichtenhofer, C., Darrell, T., & Xie, S. (2022). *ConvNeXt: Revisiting ConvNets in the Transformer Era*. CVPR. <https://arxiv.org/abs/2201.03545>
4. Wu, H., Xiao, B., Codella, N., Liu, M., Dai, X., Yuan, L., & Zhang, L. (2022). *TinyViT: Fast Pretraining Distillation for Small Vision Transformers*. ECCV. <https://arxiv.org/abs/2207.10542>
5. Keras Team. (n.d.). *Knowledge Distillation Example*. [https://keras.io/examples/vision/knowledge\\_distillation/](https://keras.io/examples/vision/knowledge_distillation/)