02 — Replication & Improvement (DenseNet121)

---

1. Objective of This Phase

After establishing a ResNet18 baseline, the next objective was to:

1. Replicate a commonly used architecture in knee OA literature
2. Compare its performance to the baseline
3. Improve the architecture through justified modifications

Professor this phase aligns directly with your requirement:

"Find related works, replicate one approach, and improve upon it."

---

2. Literature-Informed Architecture Selection

DenseNet121 was selected for replication because:

- It is widely used in medical imaging tasks.
- It has demonstrated strong performance in knee OA grading literature.
- Dense connectivity improves feature reuse and gradient flow.
- It performs well on fine-grained classification problems.

DenseNet architectures are commonly reported in KL grading studies for their ability to capture subtle texture variations in joint space narrowing and osteophyte formation.

---

3. Replication Phase — DenseNet121 @ 224×224

3.1 Model Configuration

- Backbone: DenseNet121 (ImageNet pretrained)
- Input resolution: 224×224
- Final classifier replaced with 5-class output
- Transfer learning applied

Two-stage training:

1. Train classifier head
2. Fine-tune final dense blocks

3.2 Training Setup

- Optimizer: Adam
- Initial learning rate: 1e-3 (head), 1e-4 (fine-tuning)
- Weight decay: 1e-4
- Scheduler: ReduceLROnPlateau
- Loss: CrossEntropy
- Class imbalance handled via WeightedRandomSampler

3.3 Replication Results (224 Resolution)

Test Performance:

- Accuracy ≈ 0.56
- Macro-F1 ≈ 0.58

Observation:

- Performance did not exceed ResNet baseline.
- DenseNet required stronger tuning.
- Adjacent KL confusion persisted.

This demonstrates that simply switching architecture does not guarantee improvement.

---

4. Improvement Strategy

Instead of arbitrary hyperparameter tuning, improvements were based on domain reasoning.

Radiographic KL grading relies heavily on:

- Subtle joint space narrowing
- Small osteophytes
- Fine texture differences

These features may not be well captured at 224×224 resolution.

Therefore, the improvement hypothesis was:

Increasing image resolution should allow better preservation of fine anatomical details and improve classification performance.

---

5. Final Improvement — DenseNet121 @ 320×320

5.1 Changes Introduced

- Increased input resolution to **320×320**
- Maintained DenseNet121 backbone
- Fine-tuned deeper dense blocks
- Used WeightedRandomSampler
- Adjusted batch size to accommodate GPU memory

No architectural changes beyond resolution and structured fine-tuning.

5.2 Final Model Configuration

- Backbone: DenseNet121
- Resolution: 320×320
- Optimizer: Adam
- LR (fine-tune): 1e-4
- Weight decay: 1e-4
- Scheduler: ReduceLROnPlateau
- Training epochs: 12

Hardware:

- GPU: Tesla T4

---

6. Final Results — DenseNet121 (320 Resolution)

Test Performance:

- **Accuracy ≈ 0.645**
- **Macro-F1 ≈ 0.654**

Per-class behavior:

- KL 0: Strong performance
- KL 3: High recall and precision

- KL 4: High F1 (severe OA detected reliably)
- KL 1 remains challenging
- Most confusion occurs between adjacent KL grades

This represents a clear improvement over:

- ResNet18 baseline
- DenseNet121 @224 replication

---

7. Comparative Summary

Model	**Resolution**	**Test Accuracy**	**Macro-F1**
ResNet18	224	~0.59	~0.62
DenseNet121 (replication)	224	~0.56	~0.58
DenseNet121 (improved)	320	**~0.645**	**~0.654**

Key Insight:

Resolution plays a critical role in fine-grained KL grading tasks.

---

8. Analysis of Improvement

The performance gain suggests:

- Dense connectivity benefits from higher-resolution features.
- KL grading requires fine structural details.
- Resolution scaling is more impactful than shallow architectural swaps.

The improvement was achieved without introducing complex ensembling or heavy augmentation, maintaining model interpretability and reproducibility.

---

9. Conclusion of Replication & Improvement Phase

This phase successfully demonstrates:

- Literature-informed replication
- Structured improvement
- Quantitative gain over baseline

- A defensible methodological progression

DenseNet121 at 320×320 is selected as the model used for interpretability analysis (Grad-CAM).