📊 Part 1: KD/DKD Weaknesses vs NL/NegL Improvement Opportunities (Detailed)

面向 Aspect	KD 弱點 KD Weakness	DKD 弱點 DKD Weakness	NL 改善機會 NL Improvement	NegL 改善機會 NegL Improvement	實際指標/部署信號 Real Metrics/Deployment Signal
非目標知識保留 Non-target Knowledge Retention	耦合損失抑制非目標信號，高信心樣本被減權 Coupled loss suppresses non-target signals; high-confidence samples downweighted	改善但仍只係 logit，缺乏結構 Improved but still logit-only, lacks structural info	記憶保留跨類結構，自適應動量防止被淹沒 Memory retains inter-class structure; adaptive momentum prevents drowning	在錯誤類周圍增加「排斥」，令邊界更清晰 Add repulsion around wrong classes for sharper boundaries	Minority-F1 ↑ Macro-F1 穩定 stable
訊號平衡/超參數敏感度 Signal Balance / Hyperparameter Sensitivity	無法獨立調整目標/非目標權重 Cannot independently tune target/non-target weights	α / β 敏感，容易不穩定 α / β sensitive, easily unstable	難度/一致性門動態調節，減少對固定 β 依賴 Difficulty/consistency gates dynamically adjust, reduce β dependency	控制負樣本比例，避免過懲罰 Control negative sample ratio, avoid over-penalty	梯度峰值 ↓ Gradient spikes ↓ 訓練更穩定 Training more stable
災難性遺忘 Catastrophic Forgetting	長期訓練後少數類漂移 Minority classes drift after long training	同樣存在漂移風險 Same drift risk	記憶保留早期少數類信號，減少 drift Memory retains early minority signals, reduces drift	配合 NL 保護少數類 Works with NL to protect minority classes	Minority-F1 在 epoch 10 後不退化 no degradation after epoch 10
校準問題 Calibration Issues	教師過度自信，學生跟錯 Teacher overconfident, student follows mistakes	教師偏見可能放大 Teacher bias may amplify	一致性檢查，抑制過度自信教師指導 Consistency check dampens overconfident teacher guidance	懲罰自信錯誤，降低 ECE/NLL Penalize confident mistakes, reduce ECE/NLL	ECE ↓ ≥20% NLL ↓ ≥10%
實時穩定性 Real-time Stability	輸出 jitter，翻類頻繁 Output jitter, frequent label flips	同樣缺乏 temporal consistency Same lack of temporal consistency	時序記憶平滑輸出，降低 flip rate Temporal memory smooths output, reduces flip rate	提供更安全閾值，減少錯誤觸發 Safer thresholds reduce wrong triggers	Flip rate ↓ 用戶感知穩定性 ↑ User-perceived stability ↑
Domain Shift / Noise	對 webcam noise 敏感 Sensitive to webcam noise	同樣敏感 Same sensitivity	自適應動量提升穩健性 Adaptive momentum improves robustness	合成噪聲/互補標籤訓練，提升 robustness Synthetic noise/complementary labels improve robustness	在光照/遮擋場景下表現更穩定 More stable under lighting/occlusion

Part 2: Research Framework (對應改進)

Step	Hypothesis	Method	Evidence	Reflection	Conclusion
1. Verify NL 驗證 NL	NL 防止 catastrophic forgetting NL prevents catastrophic forgetting	NL loop (短期 vs 長期) NL loop (short vs long term)	Loss 曲線、Macro-F1 隨時間穩定度 Loss curves, Macro-F1 stability over time	長期 F1 無下降 → 成立 Long-term F1 no drop → holds	NL 適合作為長期策略 NL suitable as long-term strategy
2. Verify NegL 驗證 NegL	NegL 改善 calibration NegL improves calibration	NegL loop (complementary labels)	Macro-F1、ECE、NLL	ECE/NLL 改善 → 成立 ECE/NLL improved → holds	NegL 提升可靠性 NegL enhances reliability
3. Expand Data 擴充數據	Aug + minority 擴充縮窄 domain gap Aug + minority expansion narrows domain gap	Webcam-style aug + minority data	Macro-F1、Minority-F1	Minority-F1 ↑ → 有效 Minority-F1 ↑ → effective	擴充數據保留 baseline Retain expanded data as baseline
4. Teacher 教師模型	Ensemble 提供穩定 soft labels Ensemble provides stable soft labels	RN18+EffB3 KD	Macro-F1 ≈ 0.7934 Minority-F1 ≈ 0.74	Diversity 有效但 uplift 有限 Diversity effective but limited uplift	保留 ensemble teacher Keep ensemble teacher
5. Student 學生模型	MobileNetV3 calibration 最佳 MobileNetV3 best calibration	KD baseline	Macro-F1 ≈ 0.7211 Calibration 最佳 Best calibration	F1 較低但 calibration 佳 Lower F1 but better calibration	適合作為 student Suitable as student
6. Combined Test 組合測試	NL+NegL+KD+Data 提升 F1/校準 NL+NegL+KD+Data improves F1/calibration	Full training	Macro-F1, Minority-F1, ECE, NLL	與 baseline 對比 Compare with baseline	有改善 → 方法有效 Improved → method valid
7. Observe Results 觀察結果	NL+NegL 幫助校準，即使 F1 持平 NL+NegL help calibration even if F1 flat	分析結果 Analyze results	ECE 改善但 F1 持平 ECE improved but F1 flat	支持 NL/NegL 幫助 calibration Supports NL/NegL for calibration	提升可靠性 Enhances reliability
8. Next Step 下一步	Hard-sample mining / per-class calibration	設計新策略 Design new strategy	記錄差距 Document gaps	分析差距來源 Analyze gap sources	決定是否加入新策略 Decide if adding new strategies

🎨 Part 3: Integration Summary (整合特徵)

KD/DKD 弱點總結 (KD/DKD Weakness Summary)

- **耦合抑制** (Coupled suppression): Target CE + non-target KL 混合壓抑有用信號
- **超參數敏感** (Hyperparameter sensitivity): α/β/T scaling 易導致不穩定
- **災難性遺忘** (Catastrophic forgetting): Minority class 長期訓練後漂移
- **校準差** (Poor calibration): Teacher overconfidence 傳給 student
- **實時不穩定** (Real-time instability): 輸出 jitter 頻繁翻類
- **Domain shift**: Webcam noise、光照變化敏感度高

NL 改善機會 (NL Improvement Opportunities)

- **記憶保留** (Memory retention): Associative memory 保留跨 epoch 知識
- **動量平滑** (Momentum smoothing): 自適應動量防止信號被淹沒
- **難度感知** (Difficulty awareness): Dynamic gating 調節 loss weighting
- **時序一致性** (Temporal consistency): 平滑 output trajectory 減少 jitter

NegL 改善機會 (NegL Improvement Opportunities)

- **懲罰過度自信** (Penalize overconfidence): Complementary labels 降低 ECE
- **邊界銳化** (Boundary sharpening): Repulsion loss 令 decision boundary 更清晰
- **噪聲魯棒性** (Noise robustness): 合成噪聲訓練提升 domain shift 穩健性

兩者結合 (Combined Synergy)

- 一個管「記憶」，一個管「**邊界**」，互補

  One manages "memory", one manages "boundary" — complementary

- **NL 保留知識 + NegL 懲罰錯誤 = 同時改善 F1 與 calibration**

  NL retains knowledge + NegL penalizes mistakes = improve both F1 and calibration

🎤 Part 4: Presentation Outline (簡報大綱)

1. Background (背景)

Knowledge Distillation (KD):

用 teacher 模型嘅 soft targets 去訓練 student

Use teacher model's soft targets to train student

Decoupled KD (DKD):

分開 target-class knowledge 同 non-target-class knowledge，提供更靈活嘅權重

Separate target-class and non-target-class knowledge for more flexible weighting

---

2. Problems Found in KD/DKD (發現問題)

KD 問題 (KD Problems)

- **Loss 結構耦合** (Coupled loss structure):

  Target CE 同 non-target KL 混埋一齊，壓抑咗有用嘅 teacher signal

  Target CE and non-target KL mixed together suppress useful teacher signals

- 容量 **mismatch** (Capacity mismatch):

  大 teacher 嘅 logits 太平滑，令 student underfit

  Large teacher's logits too smooth, causing student underfit

- 缺乏 **feature-level guidance**:

  只靠 logits，冇 spatial/attention alignment

  Relies only on logits, no spatial/attention alignment

DKD 問題 (DKD Problems)

- 仲係 **logit-centric**:

  冇 spatial/attention alignment

  Still no spatial/attention alignment

- **Hyperparameter 敏感** (Hyperparameter sensitivity):

  α/β 同 temperature scaling 好容易令 gradient magnitude 唔穩定

  α/β and temperature scaling easily cause gradient magnitude instability

- **Teacher calibration 問題**:

  如果 teacher 本身唔準，non-target KL 會放大錯誤

  If teacher itself inaccurate, non-target KL amplifies errors

- **Ensemble 場景下反效果**:

  當 ensemble 已經好強，多加 β 反而引入 noise

  When ensemble already strong, adding β introduces noise

---

3. New Solutions (新方案)

Nested Learning (NL)

設計 (Design):

- **Memory module** → 防止 catastrophic forgetting

  Prevents catastrophic forgetting

- **Difficulty-aware gates** → 動態調整 loss weighting

  Dynamically adjust loss weighting

- **Temporal memory smoothing** → 減少 real-time jitter

  Reduces real-time jitter

- **Consistency check** → 減低 teacher miscalibration 影響

  Reduces teacher miscalibration impact

效益 (Benefits):

- 解決 catastrophic forgetting → minority class retention

- 減少 gradient instability → 更穩定訓練 (more stable training)

- 減低 teacher miscalibration → consistency check 自動 dampen 過度自信 signal

Negative Learning (NegL)

設計 (Design):

- **Complementary labels** → 懲罰 student 過度自信嘅錯誤 prediction

  Penalize student's overconfident wrong predictions

- **Boundary repulsion loss** → 令 decision boundary 更清晰，保護 minority class

  Sharpen decision boundary, protect minority classes

- **Adaptive gating** → 只喺高 entropy/不確定樣本上應用

  Apply only on high entropy/uncertain samples

效益 (Benefits):

- 改善 calibration → 降低 ECE

- Sharpen decision boundary → minority class precision 提升 (improved)

- 增加 robustness → 減少 domain shift 影響 (reduce domain shift impact)

---

4. NL + NegL Synergy (協同效應)

結合架構 **(Combined Architecture):**

- **Memory-informed rejection:**

NL consistency score 觸發 NegL → selective apply

NL consistency score triggers NegL for selective application

- **Class-aware gating:**

Minority class 用較低 NegL ratio，避免 recall 下降

Lower NegL ratio for minority classes to avoid recall drop

- **Adaptive thresholds:**

NL smoothing + NegL calibration → 更穩定 decision boundary

More stable decision boundary

整體效益 **(Overall Benefits):**

- 同時改善 F1 + ECE → minority class 表現提升，calibration 更好

Improve both F1 and ECE; better minority class performance and calibration

- Deployment readiness → student model 更穩定，縮窄 offline-online gap

More stable student model, narrow offline-online gap

🌟 Part 5: Life Analogies (生活比喻) — 幫助教授快速理解

阿媽教仔 (Mother Teaching Child)

- 仔仔做錯 → 阿媽懲罰佢 → 就係 **NegL**

Child makes mistake → Mother punishes → This is NegL

(懲罰錯誤 prediction / Penalize wrong predictions)

- **仔仔唔想再被打 → 攞嘢記住 → 就係 NL**

  Child doesn't want to be punished again → Takes notes to remember → This is NL

  (保留記憶避免再犯 / Retain memory to avoid repeating mistakes)

老師教學生 (Teacher Teaching Student)

- **老師不停提醒「呢啲犯法唔可以做」 → NegL**

  Teacher keeps reminding "these illegal things you cannot do" → NegL

  (持續懲罰錯誤行為 / Continuously penalize wrong behavior)

- **學生因為不停被提醒 → 最終記住 → NL**

  Student, because of continuous reminders → Eventually remembers → NL

  (持續學習，保留正確知識 / Continuous learning, retain correct knowledge)

升華 (Elevation)

> 生活上成日遇到呢啲情景，所以我諗到 NL + NegL 一齊用。

> These scenarios happen frequently in life, so I thought of using NL + NegL together.

> 技術上就係「懲罰錯誤 + 保留記憶」，針對 KD/DKD 嘅固有問題，令 student model 更穩定、更準確。

> Technically it's "penalize mistakes + retain memory", targeting KD/DKD's inherent problems to make student model more stable and accurate.

Talking Points for Supervisor Meeting (教授面談要點)

開場白 (Opening)

> 「KD/DKD 喺 student training 上面有兩個痛點：容易忘記 minority class，仲會過度自信錯誤。」

> "KD/DKD in student training has two pain points: easily forgets minority classes, and becomes overconfident in mistakes."

> 「我就用生活比喻去理解：阿媽教仔，仔仔做錯就被懲罰（NegL），為咗避免再犯就記住（NL）。」

> "I use life analogies to understand: Mother teaching child, child makes mistakes and gets punished (NegL), to avoid repeating mistakes child remembers (NL)."

> 「老師教學生，成日提醒唔好犯法，學生就會記住。」

> "Teacher teaching student, continuously reminding not to break rules, student will remember."

> 「呢啲情景我生活上成日遇到，所以我諗到 NL + NegL 一齊用。」

> "These scenarios I frequently encounter in life, so I thought of using NL + NegL together."

Pipeline Framing

> 「我嘅 pipeline 係由 teacher ensemble 經 KD/DKD 去 student model。」

> "My pipeline goes from teacher ensemble through KD/DKD to student model."

> 「但 KD/DKD 有兩個痛點：容易忘記 minority class，仲會過度自信錯誤。」

> "But KD/DKD has two pain points: easily forgets minority classes, and becomes overconfident in mistakes."

> 「NL 就好似學生寫備忘錄，幫佢記住唔好再犯。」

> "NL is like student taking notes to remember not to repeat mistakes."

> 「NegL 就好似老師或阿媽懲罰錯誤，提醒佢唔好亂嚟。」

> "NegL is like teacher or mother punishing mistakes, reminding not to misbehave."

> 「兩者結合，就係『懲罰錯誤 + 保留記憶』，針對 KD/DKD 嘅弱點，令 student model 更穩定、更準確。」

> "Combining both is 'penalize mistakes + retain memory', targeting KD/DKD's weaknesses to make student model more stable and accurate."

Technical Details (if asked)

- **NL 實現** (NL Implementation):

  Meta-optimizer + associative memory module (64-dim, 2-layer)

  Difficulty gates + consistency checks

- **NegL 實現** (NegL Implementation):

  Complementary labels (teacher confusion matrix 指導)

  Boundary repulsion loss (adaptive gating)

- 預期改善 (Expected Improvement):

  Minority-F1 ↑ 1-2pp

  ECE ↓ 20-30%

  Real-time jitter ↓ (flip rate 12→8/min)

📈 Part 8: Expected Outcomes & Next Steps (預期成果與下一步)

短期目標 (Short-term Goals)

1. 完成 **NL Phase 1** (Complete NL Phase 1):

   - 解決 OOM 問題 (Solve OOM issues)

   - 啟用 AMP, gradient accumulation

   - 縮小 memory module (64→32 dim)

2. **NegL** 初步驗證 (Initial NegL Validation):

   - Teacher confusion matrix 指導 complementary labels

   - Boundary repulsion loss 實現

   - 測試 calibration 改善 (Test calibration improvement)

中期目標 (Medium-term Goals)

3. **NL + NegL** 組合測試 (Combined NL + NegL Test):

   - Memory-informed rejection

   - Class-aware gating

   - Full training (20 epochs, 3 seeds)

4. **Webcam** 數據擴充 (Webcam Data Expansion):

   - Hard sample mining

   - Targeted fine-tuning (1-3 epochs)

長期目標 (Long-term Goals)

5. 部署優化 (Deployment Optimization):

   - ONNX INT8 quantization

   - Per-class temperature scaling

- Real-time adaptive stabilization

6. 論文準備 (Paper Preparation):

   - 完整實驗對比 (Complete experimental comparison)

   - Ablation studies

   - Domain adaptation 分析

✅ **Summary Checklist for Meeting (會議檢查清單)**

- [ ] **Problem statement clear** (問題陳述清晰):

  KD/DKD 容易忘記 minority class + 過度自信錯誤

- [ ] **Life analogy prepared** (生活比喻準備):

  阿媽教仔、老師教學生

- [ ] **Technical solution framed** (技術方案框架):

  NL (記憶) + NegL (懲罰) = 針對性改善

- [ ] **Pipeline diagram ready** (流程圖準備):

  Teacher → KD/DKD → Student → NL + NegL → Synergy

- [ ] **Expected outcomes quantified** (預期成果量化):

  Minority-F1 ↑1-2pp, ECE ↓20-30%, Jitter ↓

- [ ] **Next steps defined** (下一步明確):

  完成 NL Phase 1, NegL 初步驗證，組合測試