**MIDDLE EAST TECHNICAL UNIVERSITY**

SEMESTER I EXAMINATION 2024-2025

# CENG 403 – Deep Learning - CNN Architecture Design & Transfer Learning (University Sources)

January 2025 TIME ALLOWED: 3 HOURS

---

INSTRUCTIONS TO CANDIDATES

1. This examination paper contains **SEVEN (7)** questions and comprises **TEN (10)** printed pages.
2. Answer all questions. The marks for each question are indicated at the beginning of each question.
3. Answer each question beginning on a **FRESH** page of the answer book.
4. This **IS NOT an OPEN BOOK** exam.
5. Show all mathematical derivations clearly with proper notation.
6. For architectural diagrams, draw clear and labeled components.
7. Calculate all requested parameters and show intermediate steps.
8. Explain computational complexity where requested.

**Question 1. Position-Sensitive Convolution Mathematical Analysis** (25 marks)
Based on Stanford CS231n and university computer vision course materials.

(a) Formulate position-sensitive convolution mathematically. Given input $X \in \mathbb{R}^{H \times W \times C}$, derive the augmented input formulation: (10 marks)

- Show $X_{aug}(i,j) = [X(i,j), \frac{i}{H}, \frac{j}{W}]$ where coordinates are normalized to [0,1]
- Explain why vanilla convolution fails for position estimation tasks
- Calculate the increased computational cost: memory and FLOPs for coordinate channels

(b) Analyze the mathematical foundation of pooling invariances. For max pooling with receptive field $R$ and stride $s$: (10 marks)

- Derive translation invariance: if $X'(i,j) = X(i+\delta, j+\delta)$ where $|\delta| < s$, show when $\text{MaxPool}(X') = \text{MaxPool}(X)$
- Calculate the exact translation tolerance for different pooling configurations
- Compare mathematical properties of max pooling vs. average pooling for invariance

(c) Design an experimental validation for position-sensitive convolution effectiveness: (5 marks)

- Propose synthetic datasets for controlled position estimation evaluation
- Define quantitative metrics for position accuracy assessment
- Statistical significance testing for performance comparison

**Question 2. Global Average Pooling vs. Fully Connected Layers** (28 marks)
Based on MIT 6.034 and university machine learning course materials.

(a) Analyze the parameter explosion problem in fully connected layers. For a CNN with final feature map of size $H \times W \times C$ and $N$ output classes: (12 marks)

- Calculate exact parameter count for FC approach: $(H \times W \times C) \times N + N$
- Show GAP parameter count: $C \times N + N$
- For AlexNet ($6 \times 6 \times 256 \to 4096 \to 4096 \to 1000$), compute parameter reduction percentage
- Analyze memory footprint implications for training and inference

(b) Prove the mathematical equivalence between GAP and specialized channel learning: (10 marks)

- Given feature map $F_k(x, y)$ for channel $k$, show GAP output: $g_k = \frac{1}{HW} \sum_{x,y} F_k(x, y)$
- Derive how network learns to optimize: $\max_{F_k} \sum_k w_{k,c} \cdot g_k$ for class $c$
- Explain why this forces $F_k$ to become confidence maps for specific objects
- Mathematical justification for why "one fully connected layer is sufficient"

(c) Evaluate GAP limitations and failure cases: (6 marks)

- When does the assumption $C \approx N$ (channels  classes) break down?
- Mathematical analysis of information loss compared to FC layers
- Propose hybrid architectures combining GAP benefits with FC expressiveness

**Question 3. Fully Convolutional Networks Theory and Implementation** (22 marks)

Based on UC Berkeley computer vision courses and research literature.

(a) Derive the mathematical transformation from FC to convolutional layers: (10 marks)

- For FC layer with weight matrix $W \in \mathbb{R}^{M \times N}$ and input $x \in \mathbb{R}^N$
- Show equivalence: $y = Wx \;\; y = \text{Conv}(x, W_{reshaped})$ with appropriate kernel size
- Calculate output dimensions: input $H' \times W' \to$ output $(H' - H + 1) \times (W' - W + 1)$ prediction maps
- Prove that this enables processing of arbitrary input sizes

(b) Analyze computational efficiency of fully convolutional approach: (8 marks)

- Compare computational cost: multiple forward passes vs. single FCN pass
- Calculate speedup factor for image classification on different resolution inputs
- Memory usage analysis: activation map storage vs. multiple network copies

(c) Design FCN applications beyond image classification: (4 marks)

- Semantic segmentation: pixel-level classification formulation
- Object detection: sliding window with efficiency gains
- Heatmap generation for localization tasks

**Question 4. CNN Architecture Design Principles and Empirical Analysis** (30 marks)

Based on comprehensive university deep learning course materials and research findings.

(a) Analyze the empirical study on filter size, depth, and width trade-offs: (15 marks)

- Mathematical formulation: for fixed computational budget $B$, analyze trade-off between depth $D$, width $W$, and filter size $F$
- Explain why "deeper networks with smaller filters provided better results"
- Calculate effective receptive field: for $L$ layers with filter size $F$, show receptive field $= 1 + L(F - 1)$
- Prove that depth provides exponential expressiveness increase while maintaining linear parameter growth

**Draw effective receptive field growth with depth**

**Show: 2 layers $\times$ 3$\times$3 vs. 1 layer $\times$ 5$\times$5 comparison**

(b) Design memory-efficient CNN architectures using the "reduce dimensionality early" principle: (10 marks)

- Calculate memory footprint: for input $224 \times 224 \times 3$, compare memory usage with stride=1 vs. stride=4 in first layer

- Justify why "information is redundant in earlier layers" from information theory perspective
- Design optimal stride schedule for 8-layer network balancing memory and performance

(c) Evaluate architectural design patterns across successful CNN families: (5 marks)

- Progressive resolution reduction: mathematical analysis of optimal reduction schedule
- Channel expansion strategies: when and why to increase feature map depth
- Computational vs. accuracy trade-offs in mobile architectures

**Question 5. Transfer Learning Mathematical Framework and Analysis** (25 marks)

Based on domain adaptation theory and university machine learning courses.

(a) Formalize the four transfer learning scenarios using mathematical notation: (12 marks)

Let source domain $\mathcal{D}_s = \{X_s, P(X_s)\}$, target domain $\mathcal{D}_t = \{X_t, P(X_t)\}$, source task $\mathcal{T}_s = \{Y_s, f_s(\cdot)\}$, target task $\mathcal{T}_t = \{Y_t, f_t(\cdot)\}$:

- Scenario 1: $\mathcal{T}_s \approx \mathcal{T}_t$, $|D_t| <$ threshold $\rightarrow$ Freeze weights $\theta_{1:L-1}$, train only $\theta_L$
- Scenario 2: $\mathcal{T}_s \approx \mathcal{T}_t$, $|D_t| >$ threshold $\rightarrow$ Fine-tune all $\theta_{1:L}$ with small learning rate
- Scenario 3: $\mathcal{T}_s \not\approx \mathcal{T}_t$, $|D_t| <$ threshold $\rightarrow$ Use $\theta_{1:k}$, retrain $\theta_{k+1:L}$
- Scenario 4: $\mathcal{T}_s \not\approx \mathcal{T}_t$, $|D_t| >$ threshold $\rightarrow$ Fine-tune all layers

(b) Analyze the theoretical foundation of layer transferability: (8 marks)

- Prove why early layers learn "generic, problem-independent" features using information theory
- Mathematical justification for "later parts are problem-dependent"
- Quantify transferability: define similarity metrics between feature representations

(c) Design optimal learning rate schedules for transfer learning: (5 marks)

- Derive layer-wise learning rate adaptation: $\eta_l = \eta_0 \cdot \alpha^{L-l}$ where $\alpha < 1$
- Explain why "you can easily disrupt the learned weights" with high learning rates
- Propose adaptive learning rate methods based on layer depth and similarity metrics

**Question 6. CNN Visualization Techniques Mathematical Framework** (28 marks)

Based on interpretable AI research and university courses on explainable machine learning.

(a) Implement gradient-based saliency map generation with mathematical rigor: (12 marks)

- Derive saliency map: $S_i = \left|\frac{\partial f_c(x)}{\partial x_i}\right|$ for class $c$
- Linear approximation justification: $f(x+\epsilon) \approx f(x) + \epsilon^T \nabla_x f(x)$
- Implementation using backpropagation: chain rule application through network layers
- Compare with integrated gradients: $\text{IG}_i = (x_i - x'_i) \times \int_{\alpha=0}^{1} \frac{\partial f(x'+\alpha(x-x'))}{\partial x_i} d\alpha$

(b) Analyze occlusion-based sensitivity analysis: (10 marks)

- Formulate occlusion experiment: $\Delta_p = f(x) - f(x \odot M_p)$ where $M_p$ is occlusion mask
- Statistical significance testing for determining important regions
- Design optimal occlusion window sizes and stride patterns
- Distinguish between network memorization vs. proper feature learning

(c) Evaluate visualization quality using quantitative metrics: (6 marks)

- Localization accuracy: IoU with ground truth bounding boxes
- Fidelity metrics: correlation between saliency scores and true importance
- Computational efficiency: runtime complexity analysis for different methods

**Question 7. Advanced CNN Topics Integration and Analysis** (22 marks)

Based on comprehensive university deep learning curricula and research literature.

(a) Design a comprehensive CNN architecture combining all discussed techniques: (12 marks)

- Architecture specification: incorporate position-sensitive conv, GAP, transfer learning capability
- Mathematical analysis: parameter count, memory footprint, computational complexity
- Training strategy: multi-stage training with different learning rates for different components
- Evaluation protocol: metrics for both accuracy and interpretability

**Design comprehensive CNN architecture**

**Include: Position-sensitive conv, GAP, transfer learning, visualization**

(b) Analyze failure modes and limitations of discussed techniques: (10 marks)

- Position-sensitive convolution: computational overhead and when it's unnecessary
- GAP limitations: information bottleneck and class imbalance effects
- Transfer learning: negative transfer and domain shift problems
- Visualization techniques: interpretation biases and validation challenges

**END OF PAPER**