

Компьютерное зрение
Весна 2025 / Университет Иннополис
Промежуточный экзамен
11 марта 2025 г.
Продолжительность: 45 минут

ФИО: _____
Группа: _____

Промежуточный экзамен содержит 4 страниц (включая титульный лист) и 10 вопросов с множественным выбором и свободными ответами по различным темам курса. Вопросы различаются по сложности (4, 6 или 9 баллов за вопрос). Проверьте, нет ли пропущенных страниц. Подпишите каждый лист.

- Напишите свое имя и название группы четко и разборчиво на листе ответов.
- Внимательно прочитайте каждый вопрос и выберите один или несколько вариантов ответа, либо дайте подробный ответ, если это необходимо.
- Не разговаривайте и не общайтесь с другими студентами во время экзамена, а также не используйте электронные устройства, книги или подготовленные ответы.

За неправильные ответы не предусмотрено штрафа, поэтому рекомендуется не оставлять ни одного вопроса без ответа. После завершения работы оставьте свой экзаменационный лист в ящике, предназначенном для вашей группы. Обратите внимание, что вы можете покинуть аудиторию только после сдачи работы. Удачи!

Total Score	Grade

Question	A	B	C	D	Max. Score	Score
1					0.2	
2					0.2	
3					0.2	
4	-	-	-	-	0.5	
5	-	-	-	-	0.5	
6	-	-	-	-	1.0	
7	-	-	-	-	1.5	
8	-	-	-	-	1.7	
9	-	-	-	-	1.7	
10	-	-	-	-	2.5	

1. Какова основная функция механизма self-attention в Vision Transformers?
 - A. Он позволяет учитывать глобальные зависимости между патчами изображения.
 - B. Он заменяет необходимость использования позиционных эмбеддингов.
 - C. Он уменьшает размерность входных данных без потери информации.
 - D. Он обеспечивает локальную фильтрацию, подобно свёрточным ядрам.
2. Какую функцию потерь чаще всего применяют для борьбы с дисбалансом классов в задачах сегментации?
 - A. Кросс-энтропию.
 - B. Dice loss.
 - C. MSE (среднеквадратичная ошибка).
 - D. Hinge loss.

3. Функция потерь Triplet Loss содержит оператор \max в выражении

$$\mathcal{L} = \max \left(\|f(a) - f(p)\|^2 - \|f(a) - f(n)\|^2 + \alpha, 0 \right).$$

Почему эта функция потерь пригодна для оптимизации методами градиентного спуска, несмотря на то, что функция $\max(x, 0)$ не является дифференцируемой в точке $x = 0$?

- A. В точке $x = 0$ используется понятие субградиента, позволяющее вычислять направление градиентного спуска даже при недифференцируемости.
 - B. Недифференцируемость в одной точке не оказывает влияния на оптимизацию, так как эта точка никогда не встречается на практике.
 - C. Современные оптимизаторы основаны на численных методах, которые не требуют аналитической дифференцируемости.
 - D. Функция $\max(x, 0)$ считается дифференцируемой, так как градиент в точке $x = 0$ задаётся равным 1.
4. Дано изображение $x \in \mathbb{R}^{224 \times 224 \times 3}$, которое разбивается на патчи размером 16×16 . Рассчитайте:
 - a) Количество патчей N .
 - b) Размерность вектора каждого патча после выравнивания.

5. При фьюжне данных для оценки глубины используют взвешенное объединение данных с LiDAR и RGB-камеры. Если LiDAR-оценка глубины $D_{\text{LiDAR}} = 3.5$ м, а оценка по RGB $D_{\text{RGB}} = 4.0$ м, и веса фьюжена равны $w_1 = 0.6$ и $w_2 = 0.4$ соответственно, вычислите итоговую глубину D_{final}

6. Даны Ground Truth маска (100 пикселей объекта, 900 фона), предсказание 80 TP (True Positives), 20 FN (False Negatives), 50 FP (False Positives). Рассчитайте IoU (отношение пересечения к объединению), Dice Coefficient (отношение удвоенного пересечения к общей площади). Учтите, что метрики используются для оценки качества сегментации объекта, а не фона.

7. Опишите разницу между использованием одного большого свёрточного ядра (например, 7×7) и последовательным применением нескольких малых ядер (например, трёх последовательных свёрток 3×3) в архитектуре CNN. Охарактеризуйте влияние на: эффективное поле зрения, вычислительную стоимость, возможность введения нелинейностей.

8. Рассмотрим сверточный слой, получающий входной тензор размером $32 \times 32 \times 3$. Слой применяет 64 фильтра размером 5×5 с шагом (stride) $s = 2$ и заполнением (padding) $p = 2$. 1. Вычислите пространственные размеры (высота и ширина) выходного тензора, используя формулу:

$$r = \left\lfloor \frac{n + 2p - k}{s} \right\rfloor + 1.$$

2. Определите общее число обучаемых параметров (веса и смещения) в этом слое.

9. Остаточные связи (residual connections, или skip connections) играют ключевую роль в обучении очень глубоких нейронных сетей. Объясните, как остаточные связи помогают смягчить проблему затухания градиентов. Приведите краткое математическое объяснение базового остаточного блока и обсудите, как тождественное отображение (identity mapping) способствует более стабильному обучению и повышению производительности глубоких CNN.

10. Даны два патча (токена), преобразованные в эмбединги, включающие по две части:

$$\mathbf{z}_1 = \mathbf{x}_1 + \mathbf{p}_1, \quad \mathbf{z}_2 = \mathbf{x}_2 + \mathbf{p}_2,$$

где: $\mathbf{x}_1 = [1, 0]$, $\mathbf{x}_2 = [0, 1]$ - содержательные эмбединги, $\mathbf{p}_1 = [\alpha, \beta]$, $\mathbf{p}_2 = [\gamma, \delta]$ - позиционные эмбединги.

Весовые матрицы и матрицы Q , K , V для входной последовательности $X \in \mathbb{R}^{N \times d}$ (где N количество патчей, d размерность эмбединга):

$$W_Q = W_K = W_V = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad Q = XW_Q, \quad K = XW_K, \quad V = XW_V.$$

Упрощенная формула внимания (scaled dot-product):

$$\text{Attention} = \frac{QK^T}{\sqrt{d}}.$$

1. **Без позиционных эмбедингов** ($\alpha = \beta = \gamma = \delta = 0$). Покажите, что механизм Attention делает модель нечувствительной к порядку токенов.

2. **С позиционными эмбедингами** ($\alpha \neq \gamma$, $\beta \neq \delta$). Покажите обратное, что механизм Attention чувствителен к порядку токенов.

3. **Анализ:** а) Какие задачи (NLP/CV) сильнее страдают от отсутствия позиционных эмбедингов? б) Предложите альтернативу синусоидальным эмбедингам для кодирования позиций.