Секция 4. НЕЙРОСЕТЕВЫЕ ТЕХНОЛОГИИ И ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ

ИНЖЕНЕРИЯ НАДЕЖНОГО РАСПОЗНАВАНИЯ ОБРАЗОВ

А.М. Агаев
Научный руководитель – Алексеев В.Ф., к.т.н., доцент
Белорусский государственный университет информатики и радиоэлектроники

Инженерия надежного распознавания образов (Robust Pattern Recognition) – это направление, фокусирующееся на создании моделей и систем, сохраняющих высокую точность и надежность в условиях неидеальных данных, шума, искажений, а также при преднамеренных атаках. Актуальность темы обусловлена необходимостью развертывания систем компьютерного зрения и анализа данных в реальных, непредсказуемых условиях.

В [1] содержатся фундаментальные основы построения моделей распознавания, включая архитектуры сверточных нейронных сетей (CNN), что является базой для понимания их уязвимостей и путей повышения надежности. [2] одна из первых работ, привлекшая широкое внимание к проблеме состязательных атак – целенаправленных малозаметных возмущений, обманывающих нейронные сети. Заложила основу для исследований в области надежности. Знаковой является работа [3], в которой представлен мощный метод состязательного обучения (Adversarial Training) как основного подхода к повышению устойчивости моделей распознавания образов к вредоносным атакам. Авторы [4] представляют наборы данных (ImageNet-C, CIFAR-10-C) для тестирования устойчивости моделей к естественным искажениям (шум, размытие, погодные эффекты), что является критически важным аспектом инженерии надежных систем. Подходы к интерпретируемости и отказоустойчивости как элементам надежности достаточно хорошо представлены в [5].

В докладе рассматриваются ключевые методологические подходы, влияние качества данных, вопросы устойчивости и практические рекомендации для повышения точности и надежности систем распознавания образов. В научном контексте задача формулируется как построение модели, способной аппроксимировать неизвестное отображение от визуального сигнала к понятным человеку при приеме на вход распределений, отличающихся от обучающих.

Современные системы опираются на два взаимодополняющих класса приемов. Первый класс – это архитектурные решения и алгоритмы обучения. Сверточные сети (CNN) [6] остаются базовой архитектурой для извлечения локальных инвариантных признаков. Трансформерные блоки и гибридные схемы вводят механизм глобального контекста, демонстрируют преимущество на задачах, где важна взаимосвязь между удаленными фрагментами изображения. На уровне обучения важнейшую роль играют стратегии регуляризации, схемы оптимизации и методы предобучения, позволяющие моделям получать качественные представления даже при ограниченном числе примеров. Второй класс – методы работы с данными обучающими процессорами – методы аугментации, балансировки классов, самоконтролируемого предобучения и semi-supervised подходы. В современных системах предобучение на больших неразмеченных массивах с последующим тонким

дообучением [7] на целевой задаче зачастую дает более устойчивый рост точности, чем попытки улучшения архитектуры при том же объеме данных.

Качество входных данных определяет потолок возможной точности модели. Нерегулярности разметки, смещение распределения и скрытые корреляции приводят к обобщениям и систематическим ошибкам. Важно ориентироваться на правильные источники, регламентировать доступ к базе данных, обеспечивать фильтрацию данных и выполнять аудит новых данных и качества разметки и метрики для оценки информационной полезности каждого примера. В практическом применении зачастую эффективнее улучшить разметку и покрытие кейсов, чем увеличивать размер или сложность модели - небольшие целевые инвестиции в качество данных и контрольные процедуры чаще дают больший эффект, чем масштабные изменения архитектуры.

Библиографический список

1. Goodfellow I. Deep Learning / I. Goodfellow, Y. Bengio, A. Courville. – Cambridge, Massachusetts: The MIT Press, 2016. – 800 p.

2. Szegedy C. Intriguing properties of neural networks [Электронный ресурс] / C. Szegedy, W. Zaremba, I. Sutskever [и др.] // arXiv preprint arXiv:1312.6199. – 2014. – URL: https://arxiv.org/abs/1312.6199 (дата обращения: 22.10.2025).

3. Madry A. Towards Deep Learning Models Resistant to Adversarial Attacks [Электронный ресурс] / A. Madry, A. Makelov, L. Schmidt [и др.] // arXiv preprint arXiv:1706.06083. – 2018. – URL: https://arxiv.org/abs/1706.06083 (дата обращения: 22.10.2025).

4. Hendrycks D. Benchmarking Neural Network Robustness to Common Corruptions and Perturbations [Электронный ресурс] / D. Hendrycks, T. Dietterich // arXiv preprint arXiv:1903.12261. – 2019. – URL: https://arxiv.org/abs/1903.12261 (дата обращения: 22.10.20255).

5. Molnar C. Interpretable Machine Learning: A Guide for Making Black Box Models Explainable. – 2022.

6. Smith B. Convolutional Neural Networks in Python: Master Data Science and Deep Learning with Modern Neural Networks written in Python and Theano [Электронный ресурс] / B. Smith. – Электрон. дан. – 2016. – URL: https://example.com (дата обращения: 22.10.2025).

7. Sutton R. S. Reinforcement Learning: An Introduction / R. S. Sutton, A. G. Barto. – 2nd ed. – Cambridge, Massachusetts: The MIT Press, 2018. – 552 p.