

УДК 004.932.2

doi: 10.15622/rcai.2025.050

РЕГРЕССИОННАЯ НЕЙРОННАЯ СЕТЬ ДЛЯ ОЦЕНКИ ПОЛОЖЕНИЯ БПЛА НА ОСНОВЕ РЕГУЛЯРИЗАЦИИ ФУНКЦИИ ПОТЕРЬ И ВЕСОВЫХ КОЭФФИЦИЕНТОВ¹

Я.С. Пикалёв (*i@pikaliov.ru*)

Б.В. Павленко (*bogdanpavl2000@mail.ru*)

Институт проблем искусственного интеллекта, Донецк

В работе представлена архитектура регрессионной нейронной сети RegrUAV, разработанной для точного определения углов поворота камеры беспилотного летательного аппарата (БПЛА) по входным аэрофотоснимкам. Предложена модифицированная функция потерь Focal Huber Loss (FHL), объединяющая свойства Huber Loss и Focal Loss, что позволяет повысить робастность модели, а также компенсировать дисбаланс данных. Для улучшения качества регрессии интегрирован метод Feature Distribution Smoothing (FDS), обеспечивающий сглаживание распределения признаков и калибровку статистики в процессе обучения. Проведено тестирование на наборах данных GTA-UAV и VisLoc с использованием различных архитектур Backbone (StripNet-small, InceptionNext-tiny). Результаты показали, что применение FHL и FDS обеспечивает снижение ошибки MAE и повышение устойчивости модели к дисбалансу и вариативности данных, при этом на данных GTA-UAV архитектура StripNet-small продемонстрировала наилучшие показатели точности.

Ключевые слова: свёрточные нейронные сети, регрессия, БПЛА.

Введение

Определение угла поворота камеры беспилотного летательного аппарата (БПЛА) на основе входного изображения представляет собой важную задачу в области компьютерного зрения и обработки изображений [Пикалёв, 2024]. Эта задача имеет множество направлений применения и значительно влияет на эффективность работы БПЛА в различных сферах [Степанов, 2011].

¹ Работа выполнена в рамках федерального проекта «Развитие человеческого капитала в интересах регионов, отраслей и сектора исследований и разработок» национального проекта «Наука и университеты».

Угол наклона камеры (рис. 1) влияет на интерпретацию сцены, что критично для задач распознавания объектов, сбора данных и анализа изображений. Использование нейронных сетей и методов регрессии для определения угла поворота камеры помогает в создании более точных моделей, обученных на больших объемах данных, что улучшает их производительность, позволяя адаптировать действия БПЛА в реальном времени.

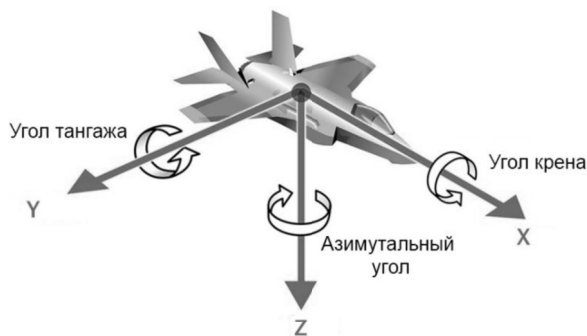


Рис. 1. Схема поворота углов БПЛА

Цель данной работы заключается в разработке и реализации модели, способной точно определять углы поворота камеры БПЛА на основе входных изображений, захваченных во время полета.

Стоит отметить, что отдельной проблемой является дисбаланс данных [Павленко, 2024]. Вместо сохранения идеального равномерного распределения по каждой категории данные часто демонстрируют искаженное распределение с длинным хвостом, когда для определенных целевых значений наблюдается значительно меньше наблюдений. Это явление ухудшает качество моделей глубокого обучения.

Таким образом, в работе были поставлены следующие задачи.

1. Спроектировать архитектуру нейронной сети RegrUAV, состоящую из таких составных частей, как Backbone, Neck, Head.
2. Изменить процесс обучения нейронной сети, используя технику сглаживания распределения признаков (Feature Distribution Smoothing, FDS).
3. Разработать функцию потерь Focal Huber Loss (FHL), учитывающую дисбаланс данных.

Провести численные эксперименты и сравнить эффективность применения FDS и FHL.

1. Архитектура RegrUAV

1.1. Общая схема

На рис. 2 изображена общая схема RegrUAV, состоящая из:

- 1) backbone – модуля, который извлекает начальные признаки из входных изображений;
- 2) neck – модуль, который манипулирует и объединяет признаки, извлечённые из backbone;
- 3) head – модуль, который преобразует признаки в финальный выход (значения углов ориентации).

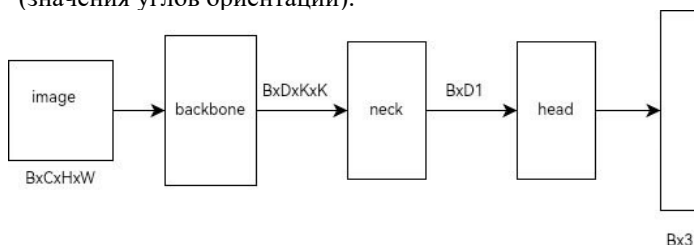


Рис. 2. Общая схема нейросети для определения положения углов наклона БПЛА (image – входное изображение, В – размер батча, С – количество каналов, Н – высота изображения, W – ширина, D – размер вектора визуальных признаков, К – количество ядер, D1 – размер одномерного тензора)

1.2. Backbone

Модуль Backbone является важнейшим компонентом в архитектурах нейронных сетей, преобразуя входные данные в сжатое скрытое представление и фиксируя существенные признаки, отбрасывая нерелевантные детали относительно поставленных целей. Предлагаемый backbone основан на модели StripNet (конфигурации StripNet-S) [Yuan, 2025].

Объекты небольшого масштаба довольно часто встречаются на аэрофотоснимках. При этом эффективность обнаружения снижается по мере увеличения масштаба объектов. В StripNet используются ортогональные крупноядерные (large-kernel) свертки в качестве основных пространственных фильтров, которые хорошо показывают себя при работе с аэрофотоснимками.

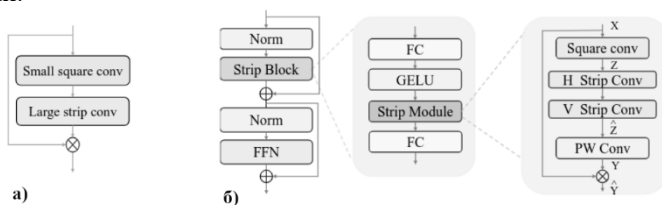


Рис. 3. Ключевые компоненты StripNet [Yuan, 2025]

StripNet состоит из базовых блоков, как показано на рис. 3,а: двух блоков с остаточной связью: strip-блока и полносвязного слоя. Как видно на рис. 3,б, strip-блок построен на стандартной свертке с малым ядром и двух свертках с большими ядрами в форме полосы (large strip convolutions) для захвата надежных признаков для объектов различного масштаба.

1.3. Neck

Модуль Neck (рис. 4) был спроектирован для захвата контекста при различных масштабах и обеспечения устойчивости (робастности) к изменениям масштаба и разрешения.

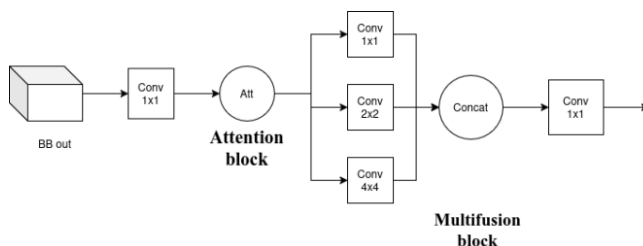


Рис. 4. Общая схема модуля Neck

Его основными блоками являются блок внимания (Attention Block, рис. 5), который улучшает точность регрессии и обеспечивает взаимодействие с разреженными слоями, улучшает различие признаков; и блок многомасштабного извлечения признаков (Multifusion Block), который объединяет признаки, собранные на разных уровнях, в единое представление, что позволяет учитывать как локальные, так и глобальные контексты.

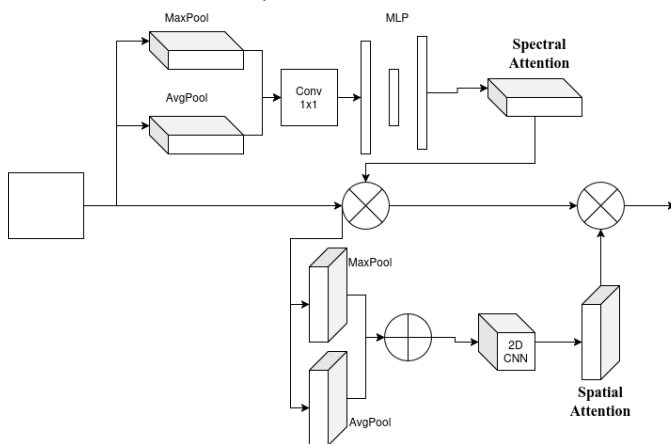


Рис. 5. Общая схема Attention Block

1.4. Head

Модуль Head (рис. 6) состоит из двух полносвязных (Fully-Connected, FC) для определения на выходе 3х углов: pitch (тангаж), yaw (рысканье), roll (крен).

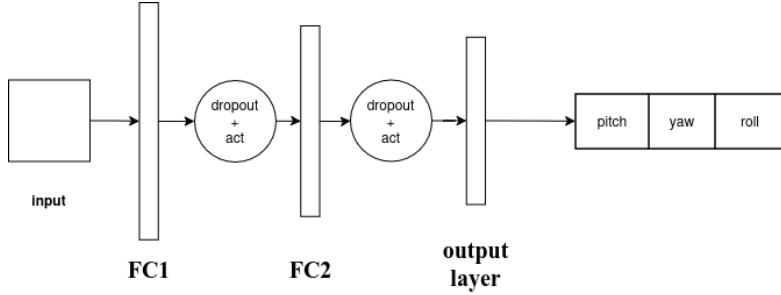


Рис. 6. Общая схема модуля Head

2. Сглаживание распределения признаков для задачи регрессии

Для решения проблемы дисбаланса данных используется FDS [Yang, 2021], которое выполняет сглаживание распределения в пространстве объектов, т.е. переносит статистику объектов между ближайшими целевыми ячейками. Эта процедура направлена на калибровку потенциально необъективных оценок распределения признаков в обучающих данных.

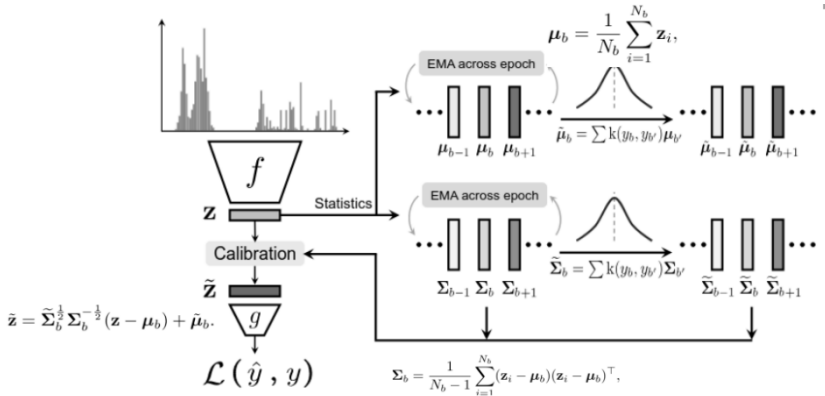


Рис. 7. Схема работы FDS в процессе обучения модели регрессии [Yang, 2021]

Для того, чтобы не ухудшить способность модели к обобщению признаков используется ковариация, чтобы отразить также взаимосвязь между различными функциональными элементами в пределах z :

$$\text{---} \quad (2.1)$$

$$\text{---} \quad (2.2)$$

где N – общее количество образцов в ячейке. Учитывая статистику признаков, используется функция сглаживания в виде симметричного ядра K , чтобы сгладить распределение среднего значения признака и ковариации по целевым ячейкам B . Это приводит к сглаженной версии статистики:

$$\hat{\mu}_B = \frac{1}{|B|} \sum_{i \in B} \mu_i \quad (2.3)$$

$$\hat{\Sigma}_B = \frac{1}{|B|} \sum_{i \in B} \Sigma_i \quad (2.4)$$

$$\hat{\mu}_B = \frac{1}{|B|} \sum_{i \in B} \mu_i \quad (2.5)$$

FDS интегрируется в RegrUAV путём добавления слоя калибровки характеристик после формирования признаков в Backbone. Для обучения модели используется динамическое обновление текущей статистики по каждой эпохе обучения. Соответственно, сглаженная статистика обновляется в разные периоды, но фиксируется в пределах каждого периода обучения.

В качестве функции обновления весов используется экспоненциальное скользящее среднее (Exponential Moving Average, EMA) [Morales-Brotons, 2024] для текущей статистики, что приводит к более стабильным и точным оценкам статистики признаков во время обучения. Затем откалиброванные признаки z передаются в регрессионную функцию и используются для вычисления потерь.

3. Focal Hubert Loss

В качестве основных функций потерь для задачи регрессии используется среднеквадратичная ошибка (Mean Square Error, MSE) и Huber Loss, отображённые на рис. 8.

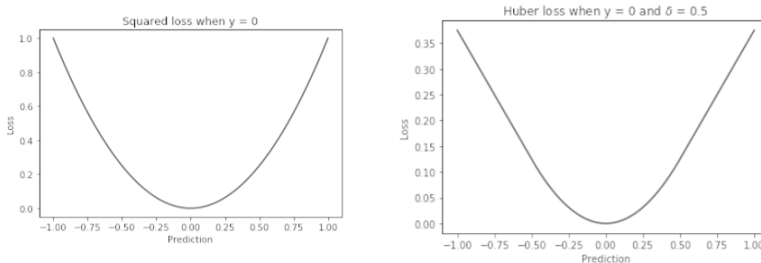


Рис. 8. Графики MSE Loss (слева) и Huber Loss (справа)

MSE «взвешивает» большие ошибки больше, чем меньшие, в зависимости от величины ошибки. Это может приводить к выбросам в случае больших значений. Одним из решений является использование Huber Loss, который менее чувствителен к выбросам, чем MSE, поскольку существует только линейная зависимость между размером ошибки и потерей за ее пределами:

$$\text{Huber Loss} = \begin{cases} \frac{1}{2}x^2 & \text{if } |x| \leq \delta \\ \delta|x| & \text{if } |x| > \delta \end{cases} \quad (3.1)$$

$$\text{Huber Loss} = \begin{cases} \frac{1}{2}x^2 & \text{if } |x| \leq \delta \\ \delta|x| & \text{if } |x| > \delta \end{cases} \quad (3.2)$$

В задаче классификации проблема дисбаланса данных (классов) решается при помощи Focal Loss. Эта функция уменьшает потери для «хорошо классифицированных примеров» или примеров, когда модель предсказывает правильные результаты с вероятностью p более 0.5. Дополнительно она увеличивает потери для «трудно классифицируемых примеров», когда модель предсказывает с вероятностью менее 0.5. Это свойство позволяет обращать «внимание» моделей на дисбаланс классов.

$$\text{Focal Loss} = -\text{CE} \times (1 - p)^{\gamma} \quad (3.3)$$

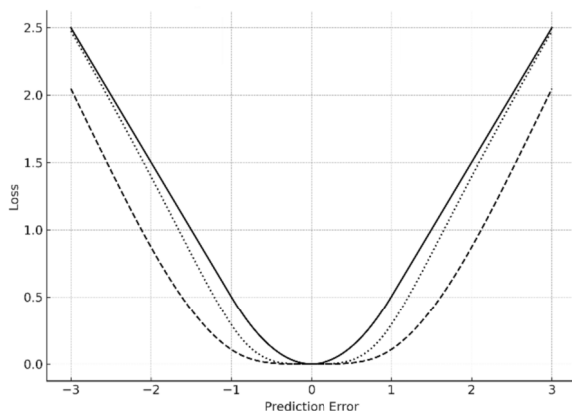


Рис. 9. Сравнение вариаций Huber Loss.

Чем выше значение γ , тем меньше потери для хорошо классифицированных примеров; α присваивает высокий вес редкому классу и низкий – распространенному.

Основная идея Focal Huber Loss (см. рис. 9) состоит в том, чтобы объединить ключевые преимущества Focal Loss с Huber Loss:

$$- \quad , \quad (3.4)$$

$$h \quad h \quad (3.5)$$

$$\text{---} \quad (3.6)$$

Главным преимуществом FHL является то, что малые ошибки имеют пониженный вес, т.е. основной акцент делается на сложные примеры, т.к. некоторые углы обладают высоким уровнем шума. Т.е. FHL решает проблему дисбаланса данных. А экземпляры данных с небольшим уровнем шума перестают оказывать сильное влияние в обновлениях градиента после того, как они изучены.

4. Характеристика обучающего и тестового набора данных

В качестве наборов данных для обучения и тестирования были выбраны GTA-UAV [Ji, 2025] и VisLoc [Xu, 2025].

GTA-UAV создан с использованием видеоигры для моделирования и сбора большого количества аэрофотоснимков, полученных с разных высот (от 80 до 650 м) и сценариев полета. В общей сложности со всей игровой карты собрано 33763 снимка с БПЛА, охватывающих различные сцены, такие как город, горы, пустыня, лес, поле и побережье. Углы БПЛА имеют следующие диапазоны:

$$\begin{aligned} \text{roll, } \phi & [-10...10]; \\ \text{pitch, } \theta & [-100...-80]; \\ \text{yaw, } \psi & [-180...180]. \end{aligned}$$

Разделение на обучающие и тестовые данные проводилось посредством объединения тестовых и обучающих данных с последующим разделением, где процент обучающих данных был равен 80% (25281 для обучающих и 8427 для тестовых).

VisLoc состоит из данных, собранных из различных регионов Китая, и содержит разные географические объекты и местности: сёла, города, фермы, реки, склоны и леса. VisLoc содержит аэрофотоснимки, полученные при разной высоте полёта (400-2000 м) и углах наклона камеры, а также различную сезонность (лето и осень). Набор данных VisLoc состоит из 6742 снимков, полученных с БПЛА (с аннотацией в виде высоты, углов поворота и географических координат) и 11 спутниковых карт.

При подготовке наборов данных углы поворота приводились к положительным значениям.

5. Численные эксперименты

В ходе численных экспериментов проводится обучение и тестирование RegrUAV на основе Focal Huber Loss и средней абсолютной ошибки (Mean Absolute Error, MAE).

В данной задаче используется MAE вместо коэффициента детерминации R^2 из-за ряда недостатков: нестабильность при малой дисперсии и чувствительность к выбросам в данных.

В качестве Backbone использовались предобученные архитектуры: StripNet-small и InceptionNext-tiny [Yu, 2024].

Обучение проводилось со следующими параметрами:

- размер входного изображения: 384x384 пикселей;
- размер батча при обучении: 64;
- размер батча при тестировании: 16;
- количество эпох: 10;
- оптимизатор: Ranger [Wright, 2021].

Таблица 1

Сравнительная таблица улучшения количественных показателей
RegrUAV

Backbone	Набор данных	test loss	train loss	test MAE	train MAE
StripNet-small	GTA-UAV	0.05	0.05	0.1	0.11
InceptionNext-tiny	GTA-UAV	0.1	0.08	0.21	0.17
StripNet-small	VisLoc	0.11	0.10	0.22	0.22
InceptionNext-tiny	VisLoc	0.11	0.11	0.21	0.24

Исходя из результатов обучения в табл.1 можно сделать вывод, что Backbone StripNet-small показал минимальные значения ошибок, что указывает на высокую способность архитектуры извлекать информативные признаки с аэрофотоснимков и эффективно обучаться на данном наборе данных.

На наборе данных VisLoc рассматриваемые архитектуры показали более высокие ошибки, чем на GTA-UAV, что говорит о большей сложности задачи и о меньшем объеме данных. Разница между StripNet-small и InceptionNext-tiny минимальна по test MAE (0.22 и 0.21), но train MAE у Inception выше (0.24 и 0.22), что может указывать на чуть большую склонность к переобучению.

Таким образом, StripNet-small демонстрирует преимущество на более специфичных для аэрофотосъёмки данных (GTA-UAV), где архитектура, оптимизированная под особенности снимков с БПЛА, лучше справляется с извлечением признаков.

InceptionNext-tiny на VisLoc показывает сопоставимые результаты по тестовой ошибке, что может объясняться тем, что этот набор содержит более разнообразные сцены, где универсальные архитектуры также работают хорошо. Ошибки на VisLoc примерно в два раза выше, чем на GTA-UAV, что указывает на необходимость дополнительной адаптации или применения методов доменной адаптации.

Из вышесказанного можно сделать вывод, что для задачи определения углов ориентации камеры на данных, подобных GTA-UAV, предпочтительно использовать StripNet-small. Для более сложных и разнообразных данных, подобных VisLoc, следует рассматривать улучшение методов регуляризации, увеличение объёма данных или комбинированные архитектуры, чтобы снизить ошибки.

Заключение

Спроектированная архитектура RegrUAV для задачи определения поворота углов по входному изображению с БПЛА обладает следующими ключевыми особенностями:

- использование backbone специализированного под данные формата аэрофотоснимков;
- использование Attention Block, который улучшает точность регрессии и обеспечивает взаимодействие с разреженными слоями, улучшает различие признаков;
- использование Multifusion Block, который объединяет признаки, собранные на разных уровнях, в единое представление, что позволяет учитывать как локальные, так и глобальные контексты;
- использование алгоритма сглаживания весовых коэффициентов после каждой эпохи обучения;
- использование функции потерь Focal Huber Loss.

RegrUAV показала на 10 эпохах следующие результаты на:

1) наборе данных GTA-UAV:

- по MAE: 0.11 на обучающих данных; 0.1 на тестовых данных;
- по FHL: 0.05 на обучающих данных; 0.05 на тестовых.

2) наборе данных VisLoc:

- по MAE: 0.22 на обучающих данных; 0.22 на тестовых данных;
- по FHL: 0.1 на обучающих данных; 0.11 на тестовых.

Список литературы

- [Пикалёв, 2024] Пикалёв Я.С. Обнаружение ключевых объектов и перекрёстная геолокализация: Анализ наборов данных и методологические перспективы // Проблемы искусственного интеллекта. – 2024. – Т. 35. – №. 4. – С. 25-37.
- [Степанов, 2011] Степанов Д.Н., Тищенко И.П. Задача моделирования полета беспилотного летательного аппарата на основе системы технического зрения // Программные системы: теория и приложения: электрон. научн. журн. – 2011. – № 4(8). – С. 33-43. – URL: http://psta.psiras.ru/read/psta2011_4_3343.pdf.

- [**Павленко, 2024**] Павленко Б.В., Пикалёв Я.С. Методика создания набора аэро-фотоснимков для задачи перекрёстной геолокализации // Проблемы искусственного интеллекта. – 2024. – Т. 35, № 4. – С. 101-112.
- [**Yuan, 2025**] Yuan X. et al. Strip R-CNN: Large Strip Convolution for Remote Sensing Object Detection. arXiv 2025 //arXiv preprint arXiv:2501.03775.
- [**Yang, 2021**] Yang Y. et al. Delving into deep imbalanced regression // International conference on machine learning. – PMLR, 2021. – P. 11842-11851.
- [**Morales-Brotons, 2024**] Morales-Brotons D., Vogels T., Hendrikx H. Exponential moving average of weights in deep learning: Dynamics and benefits // arXiv preprint arXiv:2411.18704. – 2024.
- [**Ji, 2025**] Ji Y. et al. Game4loc: A uav geo-localization benchmark from game data // Proceedings of the AAAI Conference on Artificial Intelligence. – 2025. – Vol. 39, No. 4. – P. 3913-3921.
- [**Xu, 2025**] Xu W. et al. Uav-visloc: A large-scale dataset for uav visual localization // arXiv preprint arXiv:2405.11936. – 2024.19:06
- [**Yu, 2024**] Yu W. et al. Inceptionnext: When inception meets convnext // Proceedings of the IEEE/cvf conference on computer vision and pattern recognition. – 2024. – P. 5672-5683.20:16.
- [**Wright, 2021**] Wright L., Demeure N. Ranger21: a synergistic deep learning optimizer // arXiv preprint arXiv:2106.13731. – 2021.