# СРАВНИТЕЛЬНЫЙ АНАЛИЗ НЕЙРОСЕТЕВЫХ МОДЕЛЕЙ ДЛЯ КОНТРОЛЯ НОШЕНИЯ СПЕЦИАЛЬНОЙ ОДЕЖДЫ НА РАБОЧЕМ ПЕРСОНАЛЕ С ЦЕЛЬЮ СОБЛЮДЕНИЯ ТРЕБОВАНИЙ ОТиТБ

***Курочка К.С.***
*Гомельский государственный технический университет имени П.О. Сухого*

***Аннотация.*** *Работа посвящена повышению уровня безопасности на промышленных предприятиях за счет автоматизации контроля использования средств индивидуальной защиты (СИЗ). Несоблюдение правил ношения СИЗ - ключевая причина производственного травматизма. Цель исследования - определить наиболее эффективную нейросетевую модель для детектирования СИЗ, сочетающую высокую точность и скорость. Проведен сравнительный анализ моделей YOLOv8s, YOLOv8m, Faster R-CNN+FPN и SSD300 VGG16 на специально созданном наборе данных, имитирующем реальные условия промышленного предприятия (различное освещение, ракурсы, частичное перекрытие СИЗ). Модели оценивались по метрикам mAP (mean Average Precision), FPS (кадров в секунду). В результате экспериментов модель YOLOv8 показали наилучший баланс между точностью и производительностью, превзойдя Faster R-CNN+FPN и SSD300 VGG16. Результаты демонстрируют высокую эффективность применения YOLOv8 для автоматизированного контроля СИЗ и повышения безопасности труда.*

***Ключевые слова:*** *охрана труда и техника безопасности, безопасность труда, средства индивидуальной защиты, СИЗ, производственная безопасность, мониторинг использования СИЗ, защита персонала, системы мониторинга, цифровизация безопасности, производственный травматизм, сверточные сети, YOLO, SSD, Fasten R-CNN.*

Традиционные методы контроля ношения специальной одежды, основанные на визуальном наблюдении, имеют ряд недостатков. Во-первых, они требуют значительных затрат человеческих ресурсов, так как наблюдатели должны постоянно следить за большим количеством людей. Во-вторых, человеческий фактор может приводить к ошибкам и упущениям. Наблюдатели могут уставать, отвлекаться или быть необъективными, что снижает эффективность контроля. Кроме того, не всегда возможно обеспечить непрерывный контроль, особенно в больших или опасных производственных зонах. Эти проблемы приводят к необходимости внедрения автоматизированных систем, способных быстро и точно обнаруживать людей без специальной одежды или с не полным её комплектом.

Исследование фокусируется на ключевых моделях машинного зрения: R-CNN, YOLO и SSD, которые обеспечивают точную детекцию объектов в режимах реального времени и постобработки. Данные алгоритмы находят широкое применение в медицине, биометрии, транспортной сфере и других отраслях.

К примеры различные модификации Fasten R-CNN используются медицине[1], транспорте [2]. Наиболее популярной архитектурой на рынке и среди исследователей является модели архитектуры YOLO. Модель используется в различных задачах: трекинг автотранспорта [3], поиск дефектов [4], распознавание позвонков [5] и др. задачах.

Fasten R-CNN является одним из наиболее популярных решений для задачи обнаружения объектов на изображениях и видео. Модель сочетает высокую точность и производительность благодаря использованию интегрированной сети региональных предложений. В отличие от своих предшественников, Faster R-CNN устраняет необходимость в медленных алгоритмах региональных предложений, что значительно ускоряет процесс обучения и детекции. Так же данная сеть имеет модификацию в виде Fasten R-CNN FPN. Данная модификация отличается от обычной Fasten R-CNN, наличием блока FPN (Feature Pyramid Network).

FPN - это архитектура сверточной нейронной сети, разработанная для улучшения производительности в задачах обнаружения объектов и сегментации. Основная цель FPN заключается в решении проблемы масштабирования иерархии признаков внутри сверточных нейронных сетей для улучшения качества обнаружения объектов на изображениях разного масштаба.

FPN использует иерархический подход для извлечения признаков на различных уровнях масштаба. Это достигается путем добавления в сеть дополнительных боковых сверточных ветвей, которые помогают извлекать признаки на разных уровнях детализации.

Основная идея FPN заключается в создании пирамиды признаков, где каждый уровень (или слой) представляет собой карту признаков определенного масштаба. На верхних уровнях пирамиды признаки имеют меньшее пространственное разрешение, но более высокий уровень абстракции, что полезно для обнаружения объектов больших размеров. На более низких уровнях наоборот, карты признаков имеют более высокое разрешение и меньший уровень абстракции, что полезно для обнаружения маленьких объектов.

Для улучшения передачи информации между разными уровнями пирамиды признаков FPN использует операцию объединения. Обычно это выполняется с помощью операции пулинга (например, максимального пулинга) или путем добавления дополнительных сверточных слоев. На рисунке 1 изображена архитектура данной сети.

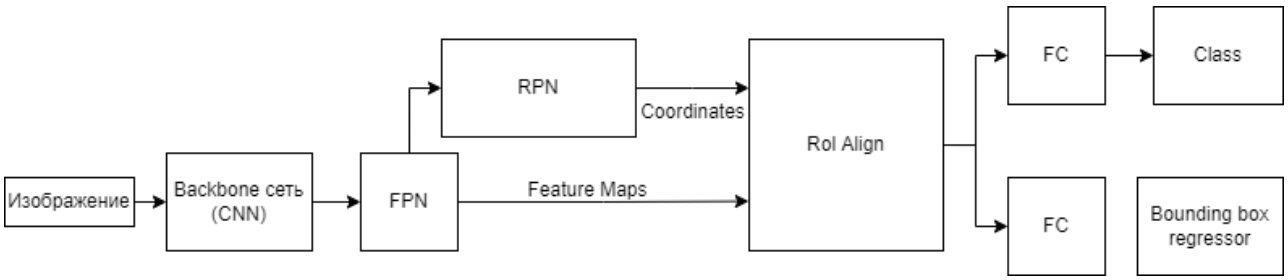


**Рисунок 1.** – Архитектура Fasten R-CNN

Также на рисунке 1 можно заметить компонент под названием backbone. Данный компонент представляет из себя свёрточную НС, которая должна генерировать карты признаков на основе входного изображения. В качестве backbone сети используется Inception

V3. InceptionV3 - это глубокая сверточная нейронная сеть, разработанная командой Google Research. Она представляет собой эволюцию изначальной архитектуры Inception, призванную улучшить как качество обнаружения объектов, так и эффективность вычислений. На рисунке 2 изображена архитектура Inception V3.

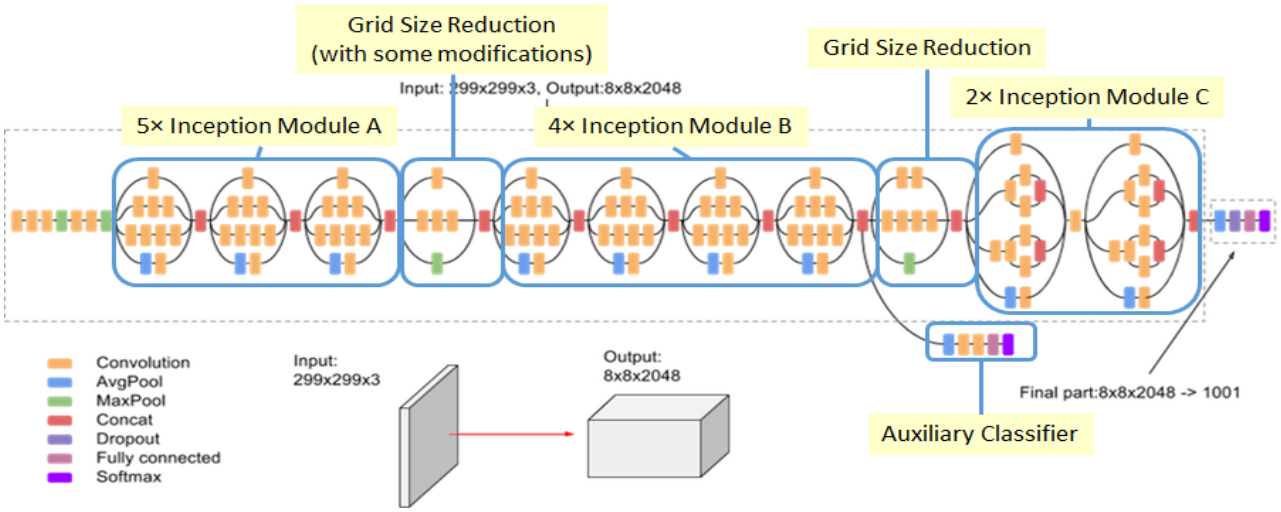


**Рисунок 2.** – Архитектура Inception V3

Глобальный пулинг (Global Average Pooling) - это операция, которая применяется к последнему сверточному слою в сети перед полносвязными слоями. Она выполняется путём вычисления среднего значения признаков по каждому каналу карты признаков. Глобальный пулинг выполняется для того, чтобы получить фиксированное количество признаков, не зависящее от размера входного изображения. Это позволяет сети работать с изображениями различного размера без необходимости использования полносвязных слоёв с фиксированным размером входа.

YOLO (You Only Look Once) - это одноэтапный детектор объектов, который делает предсказания координат, классов и уверенности одновременно. В YOLOv8 используется обновленная архитектура с оптимизированными блоками и улучшенной головой предсказания.

Все модели YOLO состоят из трёх частей: backbone, neck и head. Backbone представляется рядом сверточных слоев необходимых для извлечения базовых признаков как на ранних этапов свертки так и на поздних. Т.е. данные слои улавливают контуры и детали детектируемых объектов.

Neck служит связующим звеном между backbone и head, выполняя операции по объединению функций и интегрируя контекстную информацию. Собирается пирамида признаков, объединением карт признаков, полученных на разных этапах backbone. Данный подход повышает точность и скорость распознавания.

Головная часть (Head) нейронной сети YOLOv8 – это заключительный компонент архитектуры, который преобразует извлеченные признаки в конкретные выходные данные для задачи обнаружения объектов: модуль генерирует ограничивающие рамки, связанные с потенциальными объектами на изображении. Каждой ограничивающей рамке присваивается оценка достоверности, указывающая на вероятность присутствия объекта. На рисунке 3 изображена архитектура YOLOv8.

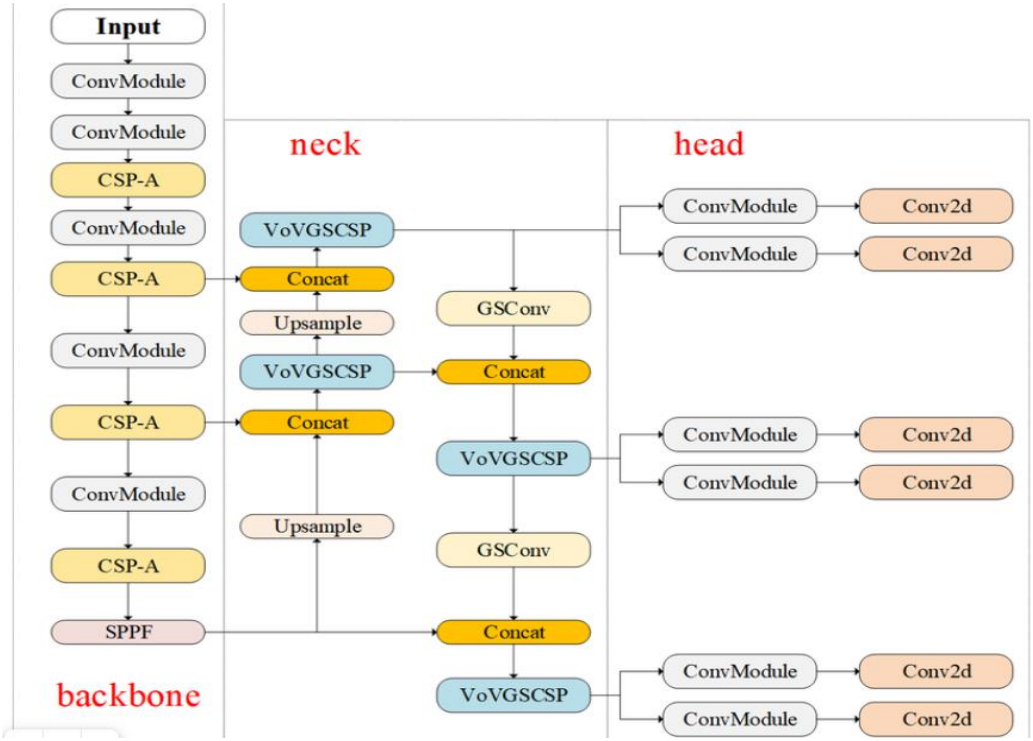


**Рисунок 3.** – Архитектура YOLOv8

Single Shot MultiBox Detector (SSD) - это одностадийная нейронная сеть, предназначенная для детекции объектов на изображениях в реальном времени. Ее ключевая особенность заключается в способности выполнять обнаружение и классификацию объектов за один проход, что обеспечивает высокую скорость работы без значительной потери точности. В оригинальной реализации SSD используется архитектура VGG16 в качестве основы для извлечения признаков из изображения. Однако, в отличие от классической VGG16, в SSD удалены полносвязные слои, что позволяет работать с изображениями переменного размера и сохранять пространственную информацию.

Сеть состоит из backbone, head и extra feature layer. Backbone это предобученная свёрточная нейросеть, которая используется для извлечения признаков из изображения. Extra feature layer необходимы для поисков признаков с разных слоёв. Данные слои повышают точность модели т.к. данный поход позволяет лучше распознавать как мелкие, так и большие объекты на разных разрешениях. Head предполагает ограничивающие рамки и принадлежность к классам. Рисунок 4 архитектура модели SSD.

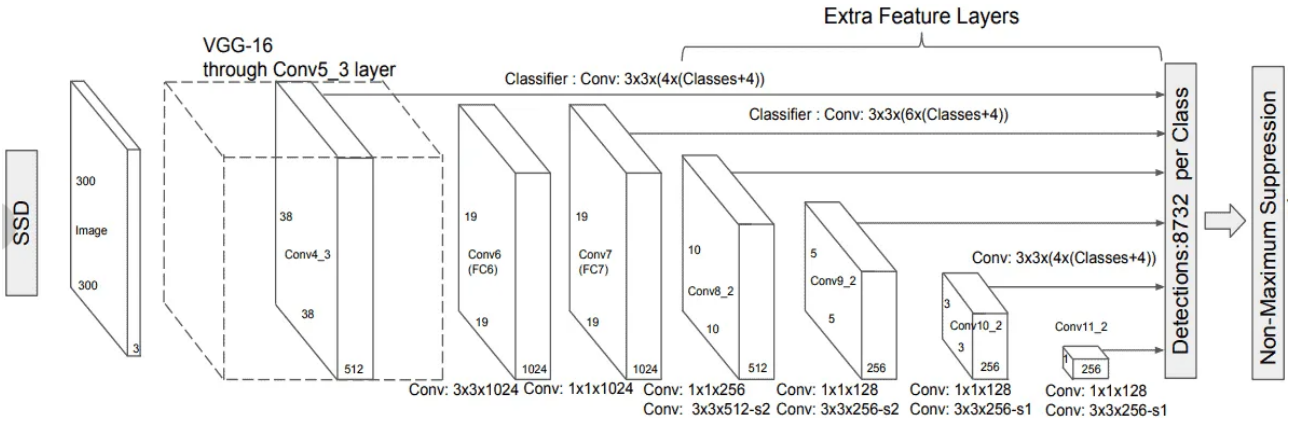


**Рисунок 4.** – Архитектура SSD

Качество обучения модели напрямую зависит от качества и объема обучающего набора данных. В данном случае создан датасет размером 900 изображений. Данные состоят из 8 классов: каска, лямка каски, спецовка, человек передом/задом, без каски, без лямки, без спецовки. В таблице 1 изображена структура данных.

**Таблица 1.** – Структура данных по колличеству объектов

| | **Каски** | **Ремешки касок** | **Спецовки** | **Отсутствие касок** | **Отсутствие ремешка каски** | **Отсутствие спецовки** | **Человек спиной** | **Человек лицом** |
|---|---|---|---|---|---|---|---|---|
| **Кол-во** | 838 | 337 | 941 | 1823 | 1609 | 1729 | 1356 | 1531 |

Для определения точности и полноты классификации использовались метрики в mAP и mAR. В сравнении использовались следующие модели: YOLO8 Medium и Small, SSD300 VGG16, Fasten R-CNN + FPN. В таблицах 2, 3 представлены результаты обучения, на рисунке 5 изображен пример маркирования модели.

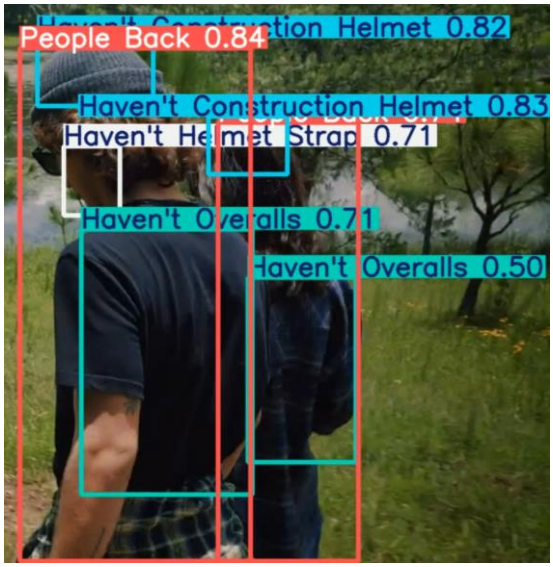


**Рисунок 5.** – Пример маркирования объектов

**Таблица 2.** – Метрика классов mAP, после обучения моделей

| **Класс** | **mAP50/FRCNN** | **mAP50/YOLO8m** | **mAP50/YOLO8s** | **mAP50/SSD** |
|---|---|---|---|---|
| Строительная каска | 0.9886 | 0.994 | 0.978 | 0.9799 |
| Нет строительной каски | 0.9898 | 0.994 | 0.992 | 0.9810 |
| Нет лямки | 0.9990 | 0.99 | 0.976 | 0.9307 |
| Нет спецовки | 0.9894 | 0.994 | 0.990 | 0.0506 |
| Лямка каски | 0.8318 | 0.948 | 0.891 | 0.9234 |
| Спецовка | 0.9898 | 0.995 | 0.986 | 0.5441 |
| Человек передом | 0.9870 | 0.989 | 0.981 | 0.1020 |
| Человек спиной | 0.9779 | 0.995 | 0.983 | 0.0463 |

**Таблица 3.** – Метрика классов mAR, после обучения моделей

| Класс | mAR50/FRCNN | mAR50/YOLO8m | mAR50/YOLO8s | mAR50/SSD |
|---|---|---|---|---|
| Строительная каска | 0.9988 | 0.997 | 0.977 | 0.9857 |
| Нет строительной каски | 0.9989 | 0.988 | 0.990 | 0.9967 |
| Нет лямки | 1.0000 | 0.993 | 0.994 | 0.9335 |
| Нет спецовки | 0.9948 | 0.995 | 0.995 | 0.0654 |
| Лямка каски | 0.8932 | 0.908 | 0.703 | 0.9318 |
| Спецовка | 0.9968 | 0.995 | 0.987 | 0.5441 |
| Человек лицом | 0.9956 | 0.979 | 0.973 | 0.2065 |
| Человек спиной | 0.9876 | 0.996 | 0.997 | 0.1803 |

В задаче обнаружения объектов наиболее эффективными архитектурами оказались YOLO v8, Faster R-CNN с FPN. Модели хорошо видят как, мелкие так и большие объекты в отличии от SSD300 VGG16. В ходе модификации параметров классификационной головы SSD300 VGG16 установлено, что базовая архитектура демонстрирует высокую точность распознавания малых и среднюю точность обнаружения средних объектов, что зафиксировано в таблицах 2 и 3. При смещении детекционного фокуса на распознавание крупных и среднеразмерных объектов наблюдается существенная деградация точности идентификации малых и очень малых объектов. Многочисленные итеративные модификации параметров не позволили существенно трансформировать детекционные характеристики модели.

Что касается YOLO8 Small она обнаруживает практически все объекты, за исключением лямок касок. Это является следствие того, что лямки касок встречаются реже всего в наборе данных, а сама модель меньше всех и имеет 11.2 миллиона параметров, когда у более старшей YOLO8 Medium 25.9 миллиона параметров. Это сказывается на обобщающей способности нейронной сети при распознавании малопредставленных классов с недостаточным количеством визуальных примеров в обучающей выборке. Это говорит нам относительно низкий mAR и высокий mAR по классу лямки касок. Тут видна тенденция на переобучение т.к. полнота определенных правильно лямок касок составляет 0.7, а точность определения 0.89.

Fasten R-CNN+FPN так же не так хорошо распознает лямки касок, но это связанно с настройкой классификационной головы, а точнее с archon generator-ом. Это видно по более низкому map над mar. Модель отлично распознает объекты, только не всегда точно определяет ограничивающие рамки. Данный признак говорит, что для классификационной головы не хватает более низкоуровневых признаков для обнаружения малых объектов, т.к. для archon generator-а использовались признаки из модуля Inception C в сверточной сети Inception V3, а это высоко уровневые слои в backbone сети.

Среди всех моделей лучше всего распознает все объекты это YOLOv8 Medium. Она лишена проблем YOLOv8 Small и Fasten R-CNN. Так же, хоть YOLO8 по некоторым классам хуже, чем Fasten R-CNN, но она быстрее классифицирует объекты чем Fasten R-CNN. В таблице 4 изображена производительность в кадрах/секундах, размер батча 30.

**Таблица 4.** – Производительность моделей в кадра в секунду

| Тип данных | YOLO8M | YOLO8S | Fasten R-CNN+FPN | SSD300 VGG16 |
|---|---|---|---|---|
| FP16 | 136 кадр/сек. | 200 кадр/сек. | 40 кадр/сек. | 136 кадр/сек. |
| FP32 | 88 кадр/сек | 166 кадр/сек. | 14 кадр/сек. | 88 кадр/сек |

Такая высокая производительность была обеспечена благодаря видеокарте RTX 3060 на 12 Гб. RTX 3060 аппаратнно поддерживает FP16 точность, это позволяет быстрее обрабатывать входные данные. Как видно из таблицы 4, при батче 30 и FP32 больше всего потеряла в скорости Fasten R-CNN+FPN. Это из-за того, что, модели не хватило видеобуфера и появилась задержка ввода/вывода. Т.е. данные полностью не поместились в оперативную память графического процессора и поступили туда порционно.

Для задач мониторинга использования одежды наиболее эффективными детекционными архитектурами являются модели YOLO v8. Они демонстрируют высокую производительность и максимальную точность распознавания, что особенно критично для систем слежения в режиме реального времени. В зависимости от производительности оборудования можно выбрать более легковесную или же тяжёлую модель. Так же при выборе более легковесной модели следует учитывать, что набор обучающих данных должен быть более сбалансированный по количеству объектов в классах. Что касается Fasten R-CNN ее нужно до настроить и ее следует использовать в случае, если необходимо классифицировать большое количество объектов на одном изображении малых и больших. SSD300 VGG16 необходимо изменить частично архитектуру классификационной головы для извлечения признаков, как низкоуровневых, для малых объектов и высокоуровневых, для больших объектов.

**Литература:**

1.Kurochka K. S., Panarin K. A. An algorithm of segmentation of a human spine X-ray image with the help of Mask R-CNN neural network for the purpose of vertebrae localization // 2021 56th International Scientific Conference on Information, Communication and Energy Systems and Technologies (ICEST) – 2021. P. 55-58. DOI: 10.1109/ICEST52640.2021.9483467

2.Ren, S., He, K., Girshick, R., & Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. // IEEE Transactions on Pattern Analysis and Machine Intelligence / Ren, S., He, K., Girshick, R., & Sun, J. Faster R-CNN. – 2017. – Vol. 39. – No 6, – P. 1137-1149.

3.Zhongjie Huang, Lintao Li1, Gerd Christian Krizek, Linhao Sun. Traffic Sign and Vehicle Detection Based on Improved YOLOv8 for Autonomous Driving // 2023 IEEE 6th International Conference on Industrial Cyber-Physical Systems (ICPS) – 2023 – P. 226-232, ISSN: 2327-5227

4.Брехт Э.А., Коншина В.Н. Применение нейронной сети YOLO для распознавания дефектов // Интеллектуальные технологии на транспорте. 2022. № 2 – 2022. – № 2. – С. 41-47.

5.Kurachka K.S., Luchshava T.V., Panarin K.A. Localization of human percentages on X-ray images with use of Darknet YOLO / Doklady BGUIR. – 2018. – Vol. 113, No. 3. – P. 32-38.