

УДК 004.89:636

ІНТЕЛЕКТУАЛЬНА МАСШТАБОВАНА СИСТЕМА ОЦІНКИ ВАГИ ВРХ ДЛЯ ЗАДАЧ ТОЧНОГО ТВАРИННИЦТВА

Черкасов А.Д.

e-mail: andrii.cherkasov2@nure.ua

Науковий керівник: к.т.н, доц., Сердюк Н.М.,

Харківський національний університет радіоелектроніки, каф. ІСТ
м. Харків, Україна

Abstract. This study presents an intelligent system for non-contact cattle live weight estimation and Body Condition Score (BCS) using computer vision. The solution integrates optimized 2D and 3D deep learning models to analyze animal geometry in real-time. By leveraging a Kubernetes-based microservices architecture, the system ensures dynamic scalability and superior GPU efficiency compared to traditional monolithic deployments. Experimental results demonstrate high precision, with weight estimation errors (MAE) under 18 kg and BCS accuracy exceeding 87 %, validating the system's effectiveness for precision livestock farming.

У сучасному точному тваринництві (Precision Livestock Farming) ключовим викликом є об'єктивна оцінка стану тварин без зайвого втручання в їхнє життя. Традиційні методи зважування та візуального огляду часто суб'єктивні, трудомісткі та, що найважливіше, викликають у худоби стрес, який негативно впливає на продуктивність. Тому було розроблено інтелектуальну систему, здатну до безконтактного визначення живої ваги та оцінки вгодованості тварини за 5-бальною шкалою [1].

Щоб комп'ютер міг «бачити» тварину так само детально, як людина, було об'єднано кілька потужних нейромережових архітектур у єдиний пайплайн. Процес починається з семантичної сегментації: модель U-Net відокремлює контур тіла тварини від фону ферми. Далі вмикаються механізми детекції та регресії на базі EfficientNetB3 та ResNet18. Вони знаходять на тілі ключові точки, такі як, холку, лінію хребта, таз, ребра у форматі COCO, що дозволяє розрахувати вагу з високою точністю – середня абсолютна похибка (MAE) становить близько 18 кг.

Однак для оцінки вгодованості плоского зображення недостатньо. Щоб зрозуміти об'єм м'язової тканини та жирового прошарку, було застосовано аналіз 3D-даних. Використовуючи карти глибини та хмари точок, архітектури PointNet++ та 3D-CNN аналізують кривизну поверхні анатомічних ділянок. Навчання моделей проводилося на фреймворках PyTorch і TensorFlow з використанням експертно розмічених та аугментованих датасетів [2].

Робота в реальному часі вимагає миттєвої реакції, тому «сирі» моделі були суттєво оптимізовані під графічні процесори. Використано CUDA 12.4 та змішану точність обчислень, а також інструменти Torch-TensorRT. Це дозволило не просто запустити моделі, а й прискорити їх виконання у 3,8–5,2 рази порівняно зі стандартним PyTorch. На практиці це означає, що затримка системи не перевищує 150 мс, навіть на високому навантаженні [2].

Замість побудови громіздкого моноліту, в основу архітектури покладено гнучкий мікросервісний підхід, де кожен компонент відповідає за свою функцію. Всю систему пропонується контейнеризувати за допомогою Docker та керувати нею через оркестратор Kubernetes із застосуванням NVIDIA GPU Operator. Таке рішення дозволить динамічно розподіляти ресурси відеокарти між різними завданнями [3].

Запропонована логіка функціонування системи відображена на архітектурній схемі (рисунк 1). Потік даних з камер надходить на Edge Gateway, на базі NVIDIA Jetson, звідки через захищений HTTPS-канал передаватиметься до кластера. Вхідний контролер Ingress (NGINX) балансуватиме навантаження та спрямовуватиме запити до API Gateway [4].

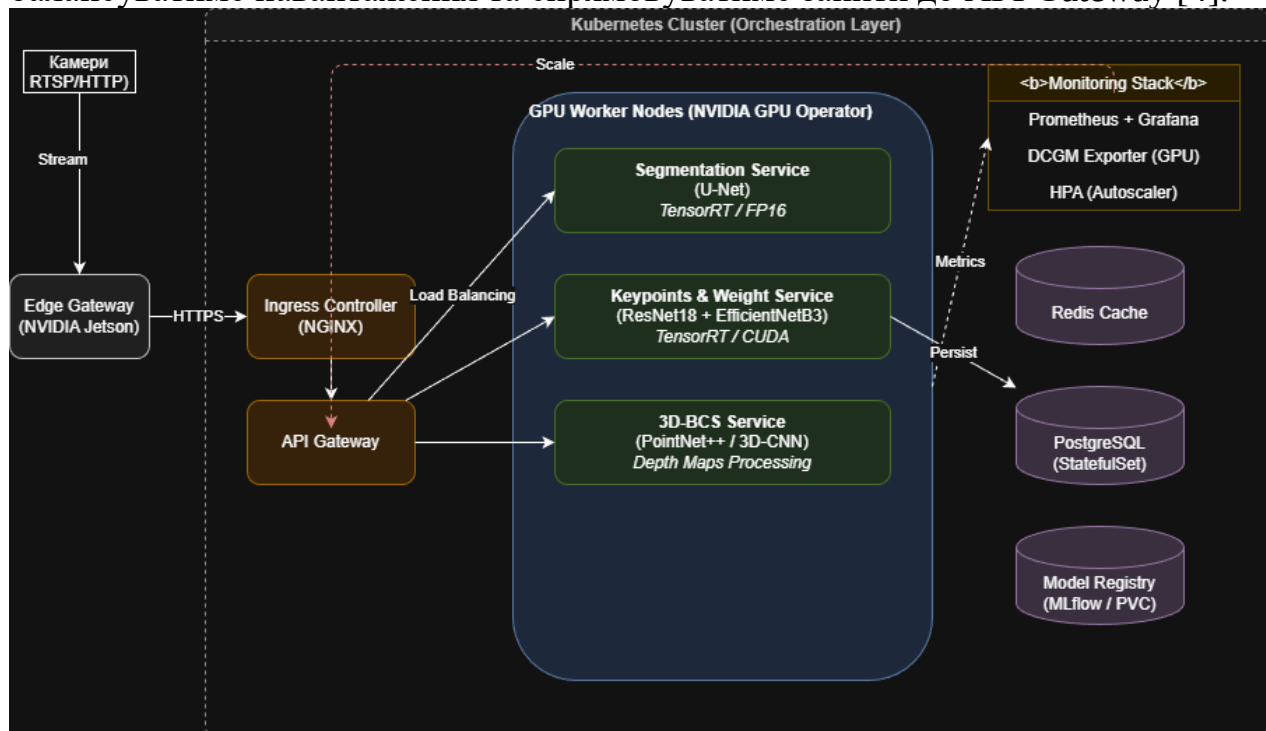


Рисунок 1 – Архітектурна схема впровадження

Серце системи – це GPU Worker Nodes, де розгорнуто три ключові мікросервіси:

- Segmentation Service (на U-Net) для виділення об'єкта;

- Keypoints & Weight Service (ResNet18 + EfficientNetB3) для розрахунку біометричних параметрів;
- 3D-BCS Service (PointNet++ / 3D-CNN) для об'ємного аналізу вгодованості.

Передбачається взаємодія цих сервісів із базами даних, Redis для кешування, PostgreSQL для зберігання стану та реєстром моделей. Важливим елементом архітектури є підсистема моніторингу на базі Prometheus та Grafana, яка у зв'язці з HPA (Horizontal Pod Autoscaler) має автоматично масштабувати кількість активних подів залежно від поточного навантаження на GPU.

Застосування такого підходу здатне кардинально змінити ефективність використання обладнання. Якщо в традиційній монолітній архітектурі відеокарти часто простоюють, то використання мікросервісів із технологією MIG (Multi-Instance GPU) дозволить підвищити їх завантаження до рівня 82–94 % [5].

Описана концепція демонструє, як поєднання глибокого навчання та сучасних DevOps-практик може трансформувати підходи у сільському господарстві. Прогнозується досягнення точності оцінки BCS понад 87 % та похибки ваги менше 18 кг у реальному часі. Використання Kubernetes дозволить не лише підвищити пропускну здатність системи на 70 %, а й зробити її надійною та легкою в обслуговуванні, що є критичним фактором для майбутнього впровадження в реальних умовах ферм.

Список використаних джерел:

1. Черкасов А. Д. Інтелектуальна система ідентифікації ваги великої рогатої худоби : кваліфікаційна робота. Харків : ХНУРЕ, 2025. 60 с.
2. Черкасов А. Д., Татарников А. О. Застосування згорткових нейронних мереж та 3D-даних для автоматизованої класифікації оцінки вгодованості великої рогатої худоби // Матеріали Молодіжної наукової ліги. Житомир, 2025. С. 315–316.
3. Hasan S. H. et al. Enhancing smart farming with containerized deep learning and Kubernetes utilizing hippopotamus optimized attention model // Malaysian Journal of Computer Science, 2025.
4. Hasan S. H., Khan A. Containerized deep learning in agriculture: Orchestrating GoogleNet with Kubernetes on high performance computing // Concurrency and Computation: Practice and Experience, 2025.
5. NVIDIA GPU Operator Documentation & livestock CV benchmarks. URL: <https://docs.nvidia.com/datacenter/cloud-native/gpu-operator> (дата звернення: 25.02.2026).