**Ковальський Семен Сергійович**

аспірант спеціальності 122 “Комп’ютерні науки”

Західноукраїнський національний університет

**Легковагові CNN для вбудованих систем контролю якості дерев'яних виробів**

У сучасних виробничих процесах дерев'яні вироби знаходять широке застосування у виготовленні меблів, декоративних елементів та конструкційних компонентів. Проте наявність дефектів (тріщини, плями, сучки) негативно впливає на якість продукції, що вимагає оперативного та точного контролю. Традиційні методи ручного огляду є затратними та схильними до суб’єктивності, що стимулює впровадження автоматизованих систем контролю якості.

З розвитком глибокого навчання Convolutional Neural Networks (CNN) стало можливим розробляти системи, здатні автоматично вилучати релевантні ознаки з зображень. Проте великі та потужні моделі, як ResNet чи DenseNet, хоч і забезпечують високу точність, мають значні обчислювальні витрати, що не дозволяє їх безпосередньо застосовувати у вбудованих системах з обмеженим ресурсом. Саме тому дослідження легковагових CNN, таких як MobileNet, EfficientNet-B0 і SqueezeNet, є надзвичайно актуальним.

Метою даної роботи є аналіз та порівняння легковагових CNN з метою визначення оптимальних архітектур для впровадження у вбудовані системи контролю якості дерев'яних виробів. Дослідження включає розробку алгоритмів попередньої обробки зображень, налаштування моделювання та оцінку продуктивності за основними метриками.

У сучасних виробничих умовах автоматизація контролю якості часто передбачає інтеграцію систем з вбудованими пристроями, що характеризуються обмеженими обчислювальними ресурсами, низьким енергоспоживанням та високою швидкістю обробки даних. Такі системи зазвичай працюють на мікроконтролерах або одноплатних комп'ютерах, де критичними є розмір моделі та час інференсу. Саме тому легковагові CNN набувають особливої актуальності для застосування у виробничих лініях.

Вбудовані системи контролю якості повинні забезпечувати оперативне виявлення дефектів навіть у режимі реального часу, не перевантажуючи обмежені апаратні засоби. Цей аспект стимулює розробку та впровадження компактних моделей, здатних працювати з мінімальними обчислювальними витратами, при цьому зберігаючи прийнятний рівень точності. Розглядаються такі легковагові моделі, як MobileNet, EfficientNet-B0 та SqueezeNet, які завдяки своїй компактності і швидкості інференсу дозволяють реалізувати ефективні системи контролю якості на вбудованих пристроях.

Для проведення експериментів було сформовано комплексний набір зображень дерев’яних панелей з дефектами та зразків без дефектів. Набір містить приблизно 2000 зображень із дефектами (дефекти представлені у вигляді тріщин, плям, сучків, смоляних кишень та інших порушень) та 1000 зразків без дефектів. Зображення було зібрано з різних джерел: внутрішніх баз даних виробництва, публічних репозиторіїв та спеціалізованих фотоколекцій, що дозволило охопити різноманіття умов освітлення, текстур і ракурсів.

Усі зображення були масштабовані до розміру 224×224 пікселів для уніфікації вхідних даних, пройшли нормалізацію та стандартизацію. Для покращення узагальнення моделей застосовано ряд технік аугментації даних, таких як випадкове обертання, горизонтальне та вертикальне віддзеркалення, зміна яскравості, контрастності і навіть невеликі зсуви по осі. Ці кроки дозволили не лише збільшити обсяг даних, а й імітувати реальні варіації зовнішнього вигляду дефектів, що є критичним для моделювання в умовах реального виробництва.

Для аналізу були обрані три легковагові CNN-архітектури: MobileNet, EfficientNet-B0 та SqueezeNet. Ці моделі були відібрані завдяки їхній компактності, низьким обчислювальним витратам та швидкості інференсу, що є важливими критеріями для вбудованих систем з обмеженими ресурсами.

Моделі навчаються із застосуванням однакових гіперпараметрів, що забезпечує коректне порівняння їх продуктивності. Зокрема, використано оптимізатор Adam із початковою швидкістю навчання 1e-4, розмір батчу встановлено на рівні 32, а кількість епох обмежено 50 з механізмом раннього зупинення, який спрацьовує при відсутності покращень на валідаційній вибірці протягом визначеного періоду. Крім того, застосовано регуляризацію за допомогою dropout і L2-регуляризації для запобігання перенавчанню моделей.

Особливу увагу приділено симуляції вбудованого середовища, оскільки однією з головних задач є розгортання моделі на пристроях з обмеженими обчислювальними ресурсами, таких як одноплатні комп’ютери з 2–4 ГБ оперативної пам’яті. Для цього моделі тестувалися не лише на потужних GPU, а й на пристроях з обмеженими ресурсами, що дозволило оцінити час інференсу (в мілісекундах) та вимоги до пам’яті. Результати цих тестів дозволяють визначити, яка з моделей краще підходить для вбудованих систем контролю якості.

Окрім стандартного розподілу даних на тренувальну, валідаційну та тестову вибірки (70/15/15), було проведено крос-валідацію для оцінки стабільності моделей та визначення оптимальних гіперпараметрів. Також аналізувалися криві навчання та валідації, що дозволяло виявити ознаки перенавчання або недонавчання і корегувати налаштування процесу.

Для оцінки якості моделей застосовувалися ключові метрики: Accuracy, Precision, Recall, F1-score та AUC (Area Under the ROC Curve). Вимірювання цих показників проводилися для позитивного класу (дефект) або як зважені середні у випадку багатокласової класифікації. Окрім цього, було проведено аналіз часу інференсу, що є важливим критерієм для систем, орієнтованих на реальний час.

Для більш детального аналізу продуктивності, результати були візуалізовані у вигляді графіків: криві навчання для кожної моделі дозволили побачити швидкість збіжності та розбіжності між тренувальною та валідаційною вибірками, а ROC-криві – їхню здатність розрізняти дефектні та бездефектні зразки при різних порогах. Також було виконано розрахунок обчислювальної складності моделей у вигляді кількості параметрів та обсягу пам’яті, що займається кожною мережею.

Основні результати представлені у таблиці 1. У дослідженні були проаналізовані три легковагові моделі: MobileNet, EfficientNet-B0 та SqueezeNet. Окрім стандартних метрик (Accuracy, Precision, Recall, F1-score та AUC), було також враховано розмір моделі (MB) та час інференсу (мс).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Модель | Accuracy | Precision | Recall | F1-score | AUC | Model Size | Inference Time |
| MobileNet | 89.7% | 91.3% | 87.2% | 89.2% | 0.920 | ~14 | 2–3 |
| EfficientNet-B0 | 91.5% | 92.0% | |  | | --- | | 90.0% |  |  | | --- | |  | | 91.0% | 0.942 | ~20 | 3–4 |
| SqueezeNet | 87.6% | 90.0% | 82.0% | |  | | --- | | 85.8% |  |  | | --- | |  | | 0.901 | ~5 | <2 |

Таблиця 1.Порівняльні результати легковагових CNN моделей для дефектування деревини.

Як видно з таблиці, EfficientNet-B0 демонструє найкращий баланс між точністю (91.5%) та обчислювальною ефективністю, забезпечуючи AUC рівень 0.942. MobileNet характеризується найменшим розміром і найшвидшим часом інференсу, проте її загальна точність дещо нижча. SqueezeNet, хоч і має найменший розмір (~5 MB), демонструє найнижчі показники точності та AUC, що свідчить про її обмежені можливості у високоточних системах контролю якості.

Під час експериментів особливу увагу приділено аналізу кривих навчання та валідації, що дозволило оцінити швидкість збіжності моделей і виявити ознаки перенавчання. Наприклад, EfficientNet-B0 демонструвала плавну збіжність з невеликою різницею між тренувальною та валідаційною точністю, що вказує на добру узагальнювальну здатність моделі. MobileNet, незважаючи на швидке навчання, показувала дещо вищу різницю між тренувальними і валідаційними показниками, а SqueezeNet характеризувалась не лише нижчими значеннями точності, а й більш вираженим розривом, що свідчить про її схильність до недонавчання.

Оскільки однією з основних цілей дослідження було розгортання моделі на пристроях з обмеженими ресурсами, проведено додатковий аналіз продуктивності на симульованому вбудованому середовищі (одноплатний комп'ютер з 2–4 ГБ оперативної пам'яті). Результати показали, що час інференсу для MobileNet складає 2–3 мс, EfficientNet-B0 – 3–4 мс, а для SqueezeNet – менше 2 мс.

Незважаючи на те, що SqueezeNet має найшвидший час обробки, її низька точність та AUC обмежують практичну користь у критичних системах контролю якості. Таким чином, EfficientNet-B0 стає оптимальним варіантом для застосування у вбудованих системах, де компроміс між продуктивністю та точністю є ключовим.

Крім часу інференсу, було проведено оцінку обчислювальної складності моделей шляхом аналізу кількості параметрів і обсягу пам'яті, що використовується. Наприклад, EfficientNet-B0 має близько 5 млн параметрів, що дозволяє значно знизити вимоги до пам’яті порівняно з класичними моделями, такими як VGG-16, яка містить понад 138 млн параметрів. Це свідчить про потенціал легковагових моделей для застосування у вбудованих системах, де обмежені ресурси є головним критерієм.

Для покращення якості моделювання проводилися експерименти з налаштування гіперпараметрів. Було протестовано кілька значень розміру батчу (16, 32, 64) та різні стратегії зміни швидкості навчання. Найкращі результати було отримано при використанні батчу розміром 32 і початкової швидкості 1e-4 із поступовим зниженням за допомогою стратегії раннього зупинення. Крім того, додаткове застосування технік аугментації, таких як випадкове обрізання зображень та регулювання контрастності, покращило узагальнювальну здатність моделей, зменшуючи ризик перенавчання.

Також проведено аналіз впливу додаткових регуляризаційних методів, зокрема застосування dropout і L2-регуляризації. Результати показали, що невеликі значення коефіцієнта dropout (0.2–0.3) сприяють стабільнішому навчальному процесу без значного зниження точності.

Загалом, результати експериментів підтверджують, що легковагові CNN мають великий потенціал для застосування у вбудованих системах контролю якості дерев'яних виробів. EfficientNet-B0 демонструє оптимальний баланс між точністю, розміром моделі та часом інференсу, що робить його найбільш придатним для умов обмежених ресурсів. MobileNet відзначається найшвидшим часом роботи, але дещо поступається за точністю, а SqueezeNet, незважаючи на свою компактність, виявляється недостатньо ефективною у високоточних завданнях.

Інтеграція додаткових оптимізацій, таких як розширена аугментація даних, оптимізація гіперпараметрів та застосування регуляризаційних методів, значно підвищує стабільність навчання та результати класифікації. Результати вказують на можливість подальшої інтеграції attention-модулів, що може ще більше покращити локалізацію дефектів на неоднорідних поверхнях деревини.

Таким чином, отримані експериментальні результати дають змогу рекомендувати EfficientNet-B0 як оптимальну модель для вбудованих систем контролю якості, тоді як MobileNet може бути використаний у випадках, де критично важлива швидкість обробки. Надалі планується розширення дослідження шляхом тестування моделей на розширених датасетах та інтеграції новітніх механізмів оптимізації для подальшого підвищення ефективності систем автоматизованого дефектування.

У цій публікації розглянуто легковагові CNN як перспективний підхід для вбудованих систем контролю якості дерев'яних виробів. Експериментальний аналіз показав, що EfficientNet-B0 забезпечує найкращий баланс між точністю, розміром моделі та швидкістю інференсу. MobileNet також є варіантом для систем з критичними обмеженнями ресурсів, хоча з дещо нижчою точністю, а SqueezeNet може бути застосована лише у випадках, коли розмір моделі є вирішальним фактором. Подальші дослідження мають зосередитися на інтеграції attention-модулів та оптимізації гіперпараметрів, що дозволить ще більше покращити продуктивність систем контролю якості.

Література

1. Howard, A. G., et al. (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv:1704.04861.
2. Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In Proceedings of the 36th International Conference on Machine Learning (ICML).
3. Iandola, F. N., et al. (2016). SqueezeNet: AlexNet-level accuracy with 50× fewer parameters and <0.5MB model size. arXiv:1602.07360.
4. Chollet, F. (2017). Xception: Deep Learning with Depthwise Separable Convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).