

УДК 004.932.2

DOI: 10.20998/2411-0558.2020.01.12

Т. О. БЛОБОРОДОВА, канд. техн. наук, доц., СХУ ім. В. Даля,
Северодонецьк,
С. О. ПРИЙМАК, магістр, СХУ ім. В. Даля, Северодонецьк,
М. О. КОВЕРГА, асп., СХУ ім. В. Даля, Северодонецьк

СУКУПНІСТЬ ЕТАПІВ РОЗРОБКИ МОДЕЛІ ВИЗНАЧЕННЯ ОБ'ЄКТІВ У СИСТЕМАХ МЕДИЧНОЇ ВІДЕОЗЙОМКИ

Представлено вирішення задачі розробки моделі визначення об'єктів у системах відеозйомки медичних процесів за рахунок встановленої сукупності етапів. В ході роботи систематизовано сукупність етапів розробки моделі, проведено експеримент з передобробки, анотації зображень відеокадрів, навчання та тестування моделі на даних лапароскопічної операції. Розглянуто основні кроки навчання та тестування моделі з використанням мережі Faster R-CNN. Точність на тестових даних визначена з використанням середньої оцінки точності моделі і дорівнює 96.3%. Іл.: 4. Бібліогр.: 12 назв.

Ключові слова: визначення об'єктів; системи медичної відеозйомки; навчання та тестування; моделі; анотація зображень відеокадрів.

Постановка проблеми. Сучасні підходи до проведення медичних процедур є орієнтованими на пацієнта і передбачають зниження заподіяння шкоди пацієнту. Це обумовлює здійснення медичних процедур через природні отвори тіла або невеликі штучні розрізи, за рахунок чого зменшується травматизм пацієнтів та скорочується час періоду реабілітації. Разом з тим, це призводить до обмеження бачення та рухливості лікаря, ускладнення координації його рук і очей, що обумовлює удосконалення засобів візуалізації, контролю та управління медичними процедурами з використанням відеозйомки. Системи відеозйомки, що використовуються при проведенні медичних досліджень, процедур або оперативних втручань, є джерелом великої кількості даних відеозображень, які можуть бути використані для підтримки прийняття клінічних рішень, виявлення патологічних діагностичних елементів, прогнозування взаємодії хірургічних інструментів і тканини внутрішніх органів, під час проведення процедури. Важливим етапом аналізу даних медичних систем відеозйомки є розробка моделі визначення об'єктів у відео. Визначення хірургічних інструментів з більш точним просторовим розташуванням в відео медичних процедур не тільки допомагає забезпечити безпеку пацієнта за рахунок зниження частоти ускладнень, але також має

значення для оцінки роботи хірурга. Існує багато факторів, що впливають на точність визначення об'єктів у відео медичних процедур: розподільна здатність відеокамери, наявність у порожнині газів, диму, запотівання лінзи камери, індивідуальні особливості анатомії тощо.

Аналіз останніх досліджень і публікацій. Визначення об'єктів – це метод комп'ютерного зору, що дозволяє ідентифікувати і визначати місце розташування об'єктів на зображенні або відео [1]. Традиційні дослідження визначення об'єктів у відео-медичних процедурах зосереджені на процесі розробки моделі визначення об'єктів у медичних відеозображеннях з використанням заздалегідь анотованих даних. Дослідження [2] проведене з використанням анотованих даних відео мікрохірургії сітківки ока. Запропонована авторами CLS архітектура згорткової нейронної мережі показала найвищу точність розпізнавання. Середній критерій точності розпізнавання становив 92.6%.

У роботі [3] автори використовують швидкі глибокі нейронні мережі з навчанням, керованим знаннями та прогнозованими областями інтересів для виявлення відеооб'єктів у режимі реального часу. Для навчання моделі використані анотовані дані. Найвищу точність отримано при використанні запропонованої мережі. Результат становив 92.37%. Автори також зазначили, що використання області інтересів дозволило скоротити час розпізнавання відео.

В дослідженні [4] запропонована модульована якірну мережу для виявлення інструментів лапароскопічної хірургії на основі Faster R-CNN, яка успадковує переваги використання області інтересів, що надає високу швидкість розпізнавання. Навчання проведено з використанням анотованих даних. Точність моделі при проведенні експериментів на двох різних відеоданих становила 69.6% та 76.5%.

В дослідженні [5] розглядається використання бінарних міток, що позначають присутність або відсутність інструменту у відеокадрі. Автори використовують ConvLSTM для навчання моделі та виявлення присутності інструменту, просторової локалізації та відстеження руху. Якість моделі визначена з використанням критерію середньої точності, що склав 92.9%.

Автори [6] запропонували підхід, який спирається на слабкі анотації, представлені у вигляді смуг поверх різних об'єктів на зображенні, і часткової крос-ентропії як функції втрат повністю згорткової нейронної мережі для отримання щільної карти прогнозування на рівні пікселів.

Найбільш успішні дослідження визначення об'єктів у відео медичних процедур використовують власноруч анотовані дані. Зазвичай, анотація зображень відеокадрів дозволяє отримати вищу точність

визначення об'єктів у відео при використанні їх для навчання моделі. Автори [7] запропонували напіваавтоматичний метод анотації відео хірургічних операцій. При застосуванні запропонованого методу необхідно анотувати невелику частину відеосегментів у необробленому відео. Результати проведеного експерименту показали, що точність визначення об'єктів залежить від відсотку анотованих сегментів відео. Точність визначена для відео з 5% та 10% анотованих сегментів. Більш точні результати отримані при 10% анотованих сегментів.

Проведений аналіз досліджень показав, що вища точність розпізнавання відео досягається за рахунок цільової анотації даних відеозображень. Також, застосування області інтересів у визначенні об'єктів дозволяє досягти зменшення часу розпізнавання. Ці факти обумовлюють актуальність поставленої мети та вибір методів її реалізації.

Мета статті. Метою дослідження є систематизація сукупності етапів розробки моделі визначення об'єктів у системах відеозйомки медичних процесів.

Основна частина. Запропонована сукупність етапів розробки моделі визначення об'єктів у системах відеозйомки медичних процесів. Вона систематизована як представлено на рис. 1.

Етапами розробки моделі визначення об'єктів у відео-медичних процесів є передобробка, анотування зображень відеокадрів, навчання та тестування моделі.

Етап 1. Передобробка відеоданих. Передобробка відеоданих включає перетворення відео на серію зображень, видалення неінформативних зображень відеокадрів та зворотнє перетворення серій зображень на файл з відео.

Етап 2. Анотація зображень відеокадрів. Анотація зображень є необхідним етапом для отримання високої точності визначення об'єктів у відео системи медичної відеозйомки. В даному дослідженні анотація об'єктів проводиться ручною розміткою досліджуваних об'єктів у відео з використанням спеціалізованого програмного забезпечення [8].

Етап 3. Навчання моделі. Для навчання моделі використовується 80% анотованих даних зображень відеокадрів. Навчання моделі проведене з використанням мережі Faster R-CNN. Навчання моделі визначення об'єктів у відео включає наступні кроки: генерація пропозицій; виявлення векторів властивостей (feature vector extraction) та класифікація областей.

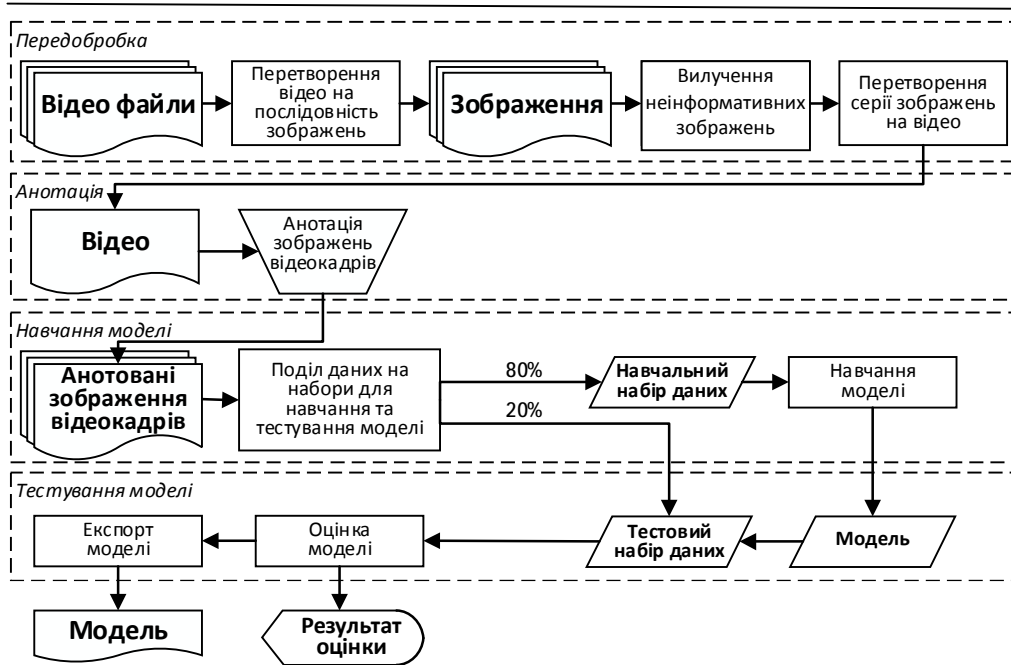


Рис. 1. Загальна сукупність етапів розробки моделі визначення об'єктів для систем відеозйомки медичних процесів

Генерація пропозицій. Під час генерації пропозицій мета полягає в пошуку областей на зображенні, які можуть містити об'єкти. Ці області називають області інтересів (regions of interest – ROI). Для того, щоб виявити інформацію про об'єкти, що мають різні розміри та співвідношення сторін, розмір вхідних зображень приводиться до різного масштабу, та використовуються ковзні вікна для просування по цим зображенням.

Виявлення векторів властивостей. Під час другого етапу на кожній позиції зображення з вікна виводиться вектор фіксованої довжини з властивостями (features) для збору семантичної інформації охопленої області.

Класифікація областей. На третьому кроці області зображення використовуються як визначені ознаки міток (labels).

Схематично архітектуру Faster R-CNN представлено на рис. 2.

Faster R-CNN має дві мережі [9]: мережу пропозицій областей (region proposal networks – RPN) для генерації пропозицій областей та мережу Fast R-CNN, що використовує ці пропозиції для визначення об'єктів.

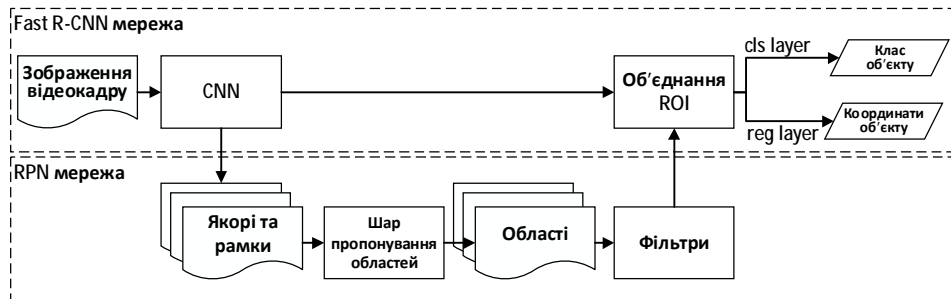


Рис. 2. Архітектура Faster R-CNN

RPN приймає зображення будь-якого розміру як вхідні дані та виводить набір пропозицій рамок з об'єктами, кожен з яких має оцінку об'єктності. Для генерації пропозицій міні-мережа пересувається по карті характеристик (conv feature map), що було отримано на виході останнього шару згорткової мережі. Кожне вікно проектується на вектор меншого виміру $256-d$, що є проміжним шаром (intermediate layer). Цей вектор подається на вхід шару регресії рамок (reg layer) та шар класифікації рамок (cls layer).

У кожному положенні вікна одночасно прогнозується k пропозицій областей, так що у шарі регресії є $4k$ виходи, що кодують координати k полів. Шар класифікації області має $2k$ виходів, які оцінюють ймовірність наявності об'єкта чи фону в області. Пропозиції зроблено відносно до k якорних регіонів (anchor boxes). Якорі відіграють важливу роль у Faster R-CNN. Кожен якор розміщений у центрі регіону, та описаний масштабом (128^2 , 256^2 або 512^2 пікселів) та відношенням сторін (1:1, 1:2, 2:1) [10]. У конфігурації Faster R-CNN за замовчуванням є 9 якорів для кожної позиції в зображенні.

Для навчання RPN кожному якорю присвоюється бінарна мітка класу – чи це об'єкт, чи ні. Позитивна мітка присвоюється двом типам якорів – якорям, що мають максимальне IoU (intersection over union) перекриття з регіоном основної правди; та якорям що мають коефіцієнт перекриття IoU з будь-яким регіоном основної правди більший ніж 0.7. Кожна область основної правди може присвоювати позитивну мітку декільком якорям. Якщо неперекриття якоря має коефіцієнт перекриття IoU для всіх областей основної правди менш ніж 0.3, то йому присвоюється негативна мітка. Якорі які не мають позитивної або негативної мітки не приймають участь в навчанні.

Функція втрат для зображення має наступний вигляд (1):

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*), \quad (1)$$

де i – це індекс якорю, а p_i – це прогнозована вірогідність того що якір i є об'єктом; мітка основної правди p_i^* має значення 1 якщо якір є позитивним, або 0 – якщо ні; t_i – це вектор з чотирьох параметризованих координат прямокутного регіону, що було спрогнозовано, а t_i^* – це прямокутний регіон основної правди, що відноситься до позитивного якорю; L_{cls} – це logloss від двох класів (чи об'єкт/чи ні). При розрахунку функції втрат для регресії використовується формула (2). Термін $p_i^* L_{reg}$ означає що функцію втрат для регресії активовано тільки для позитивних якорів, тобто $p_i^* = 1$, та виключено в іншому випадку ($p_i^* = 0$). Два терміни нормалізуються за допомогою N_{cls} і N_{reg} , та балансуєчої ваги λ

$$L_{reg}(t_i, t_i^*) = R(t_i - t_i^*), \quad (2)$$

де R – це функція втрат Хьюбера L_1 , що обчислюється за формулою (3)

$$R(x) = \begin{cases} 0.5, & \text{якщо } |x| < 1, \\ \text{інакше } |x| - 0.5. \end{cases} \quad (3)$$

Координати для регресії параметризуються за формулами (4):

$$\begin{aligned} t_x &= (x - x_a) / w_a, \\ t_y &= (y - y_a) / h_a, \\ t_w &= \log(w / w_a), \\ t_h &= \log(h / h_a), \\ t_x^* &= (x^* - x_a) / w_a, \\ t_y^* &= (y^* - y_a) / h_a, \\ t_w^* &= \log(w^* / w_a), \\ t_h^* &= \log(h^* / h_a), \end{aligned} \quad (4)$$

де x та y – це координати центру області рамки, w та h – ширина та висота області, а x , x_a та x^* (так само як і y , w , h) означають координати прогнозованої і якірної областей та області основної правди.

RPN реалізована як глибока згорткова нейронна мережа і може бути натренована за допомогою зворотного поширення та методу стохастичного градієнта. Кожна міні-партія виходить з одного зображення, яке містить багато позитивних і негативних якорів. Якщо на зображенні менше 128 позитивних зразків, то партія доповнюється негативними зразками. Всі нові шари ініціалізуються випадково, витягуючи ваги із розподілу Гауса із нульовим середнім зі стандартним відхиленням 0,01.

Етап 4. Тестування моделі. Для тестування моделі використовується 20% анотованих даних зображень відеокадрів. Тестування проводиться з використанням навченої моделі та мережі Faster R-CNN. На кроці класифікації областей за допомогою навченої моделі областям зображення присвоюються мітки. Навчена модель визначає класи об'єктів та їх можливе розташування у кадрі.

Для оцінки точності моделі використано критерій середньої точності (MAP), що визначає середню точність для набору запитів.

Результати експерименту. Експеримент по застосуванню систематизованих етапів для розробки моделі проведено з використанням відео лапароскопічної хірургічної операції, що містить наступні об'єкти: внутрішні органи черевної порожнини, медичні інструменти хірургічного комплексу DaVinci. Реалізовано наступні етапи: передобробка відео, анотація зображень відеокадрів, навчання та тестування моделі з використанням мережі Faster R-CNN. Також, проведено експеримент з визначення об'єктів у відео лапароскопічної операції з використанням камери мобільного телефону і мобільного додатку, що було розроблено у рамках цього дослідження, та розробленої моделі.

В якості об'єктів анотації визначені інструменти хірургічного комплексу DaVinci. Комплекс DaVinci має можливість роботи з наступними змінними інструментами: скальпелі (scalpel), ножиці (scissors), кліп-аплікатори (clip applier), голки (needle driver), пінцет (graspers). Для цих інструментів було створено наступні мітки: Scalpel, Scissors, ClipApplier, NeedleDriver, Graspers.

Для анотації та навчання моделі використано відео з відкритих джерел, що демонструють роботу з комплексом DaVinci [11, 12]. Для поліпшення якості даних проведено перетворення відео на серію зображень з метою подальшого видалення кадрів, що не містять цільових об'єктів розпізнавання або кадрів, що мають низьку якість зображення. Після вилучення кадрів, що не містять потрібних об'єктів та кадрів з низькою якістю зображення потрібних об'єктів, решта зображень знов перетворена на відео.

Результат анотації одного з кадрів методом обмежуючої рамки відображено на рис. 3.

Файл анотацій містить перелік усіх анотованих відеозображень, координати кожного анотованого об'єкту та його мітку. 80% цих даних використано для навчання моделі з використанням мережі Faster R-CNN. Тестування проведено з використанням 20% анотованих даних. Отримано наступні результати тестування моделі. Середня точність визначення об'єктів складає 96.3%.

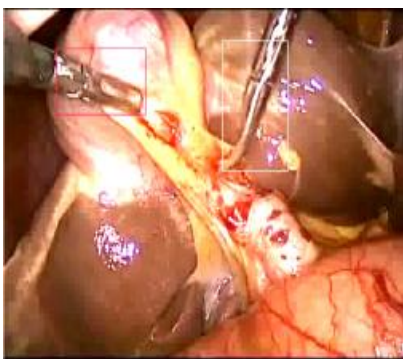


Рис. 3. Процес анотації об'єктів зображення відеокадру

Експеримент з визначення об'єктів у відео лапароскопічної операції з використанням розробленої моделі проведено в режимі реального часу. Визначення об'єктів реалізовано з використанням розробленого мобільного додатку та розробленої моделі. Додаток за допомогою камери, що вбудована в мобільний телефон, отримує відео у реальному часі, та помічає обмежуючі рамки прогнозованих об'єктів, підписує мітку об'єкта та точність його розпізнавання. Приклад процесу визначення об'єктів у відео лапароскопічної операції представлений на рис. 4.

Висновки. Метою дослідження було систематизація сукупності етапів розробки моделі визначення об'єктів у системах відеозйомки медичних процесів. Систематизовано етапи розробки моделі визначення об'єктів, що включають передобробку відео медичних процесів, анотацію зображень відеокадрів, навчання та тестування моделі з використанням мережі Faster R-CNN. Відмінною перевагою використання цієї мережі є крок генерації пропозицій. Під час генерації пропозицій мережа за допомогою ковзного вікна знаходить області на зображенні, що можуть містити об'єкти, на підставі чого формує вектор ознак, що використовується далі згортковою нейронною мережею для класифікації зображень відеокадрів.

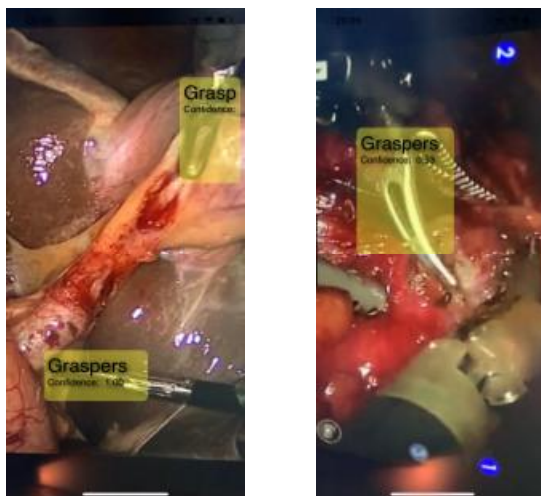


Рис. 4. Визначення об'єктів у відео лапароскопічної операції

Проведено експеримент з розробки моделі за визначеною сукупністю етапів. Середня точність визначення об'єктів на тестових даних становила 96.3%. З використанням розробленої моделі реалізовано експеримент з визначення об'єктів у відео лапароскопічної операції (рис. 4). Визначення об'єктів проведено в режимі реального часу з використанням розробленої моделі, вбудованої камери мобільного телефону та розробленого мобільного додатку.

References:

1. Kong, Y. and Fu, Y. (2018), Human action recognition and prediction: A survey. arXiv preprint arXiv:1806.11230.
2. Laina, I., Rieke, N., Rupprecht, C., Vizcaino, J. P., Eslami, A., Tombari, F., and Navab, N. (2017), "Concurrent Segmentation and Localization for Tracking of Surgical Instruments. Medical Image Computing and Computer-Assisted Intervention", *MICCAI 2017*, pp. 664-672.
3. Cao, W., Yuan, J., He, Z., Zhang, Z., and He, Z. (2018), "Fast Deep Neural Networks With Knowledge Guided Training and Predicted Regions of Interests for Real-Time Video Object Detection", *IEEE Access*, 6, pp. 8990-8999.
4. B. Zhang, S. Wang, L. Dong and P. Chen (2020), "Surgical Tools Detection Based on Modulated Anchoring Network in Laparoscopic Videos", *In IEEE Access*, vol. 8, pp. 23748-23758,.
5. Nwoye, C. I., Mutter, D., Marescaux, J., and Padoy, N. (2019), "Weakly supervised convolutional LSTM approach for tool tracking in laparoscopic videos", *International Journal of Computer Assisted Radiology and Surgery*.
6. Fuentes-Hurtado, F., Kadkhodamohammadi, A., Flouty, E. et al. (2019), "EasyLabels: weak labels for scene segmentation in laparoscopic videos", *Int J. CARS 14*, pp.1247-1257.
7. Zhiwei Fan, Congliang Chen, & Tingting Jiang (2017), "A semi-automatic editing method for surgery videos", *2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*.

8. New Computer Vision Tool Accelerates Annotation of Digital Images and Video. URL: <https://www.intel.ai/introducing-cvat/#gs.irfx8h> (access 31.03.2020)
9. Ren S, He K, Girshick R, Sun J. (2015), "Faster r-cnn: Towards real-time object detection with region proposal networks", *In Advances in neural information processing systems 2015*, pp. 91-99.
10. Lu Y., Javidi T., Lazebnik S. (2016), "Adaptive object detection using adjacency and zoom prediction", *in: CVPR*.
11. Inferior Vena Cava (IVC) Thrombectomy - da Vinci Robotic Surgery. URL: <https://www.youtube.com/watch?v=5JM8KhWhrus> (access 31.03.2020)
12. Laparoscopic Cholecystectomy. URL: <https://www.youtube.com/watch?v=n18zxNGJdLE> (access 31.03.2020)

Статью представил д.т.н., проф. Національного технічного університету "Харківський політехнічний інститут" А.А. Серков.

Надійшла (received) 31.03.2020

Biloborodova Tetiana, Cand.Sci.Tech, Associate Professor
Volodymyr Dahl East Ukrainian National University
59-a Central avenue, Severodonetsk, Luhansk region, Ukraine, 93400
tel./phone: (064) 522-89-97, e-mail: beloborodova.t@gmail.com
ORCID ID: 0000-0001-7561-7484

Prymak Serhiy, master student
Volodymyr Dahl East Ukrainian National University
59-a Central avenue, Severodonetsk, Luhansk region, Ukraine, 93400
tel./phone: (064) 522-89-97, e-mail: sprymak@gmail.com

Koverha Mark, PhD student
Volodymyr Dahl East Ukrainian National University
59-a Central avenue, Severodonetsk, Luhansk region, Ukraine, 93400
tel./phone: (064) 522-89-97, e-mail: healthunder@gmail.com

УДК 004.932.2

Сукупність етапів розробки моделі визначення об'єктів у системах медичної відеозйомки / Білобородова Т.О., Приймак С.О., Коверга М.О. // Вісник НТУ "ХПІ". Серія: Інформатика та моделювання. – Харків: НТУ "ХПІ". – 2020. – № 1 (3). – С. 143 – 153.

Представлено вирішення задачі розробки моделі визначення об'єктів у системах відеозйомки медичних процесів за рахунок встановленої сукупності етапів. В ході роботи систематизовано сукупність етапів розробки моделі, проведено експеримент з передобробки, анотації зображень відеокадрів, навчання та тестування моделі на даних лапароскопічної операції. Розглянуто основні кроки навчання та тестування моделі з використанням мережі Faster R-CNN. Точність на тестових даних визначена з використанням середньої оцінки точності моделі і дорівнює 96.3%. Іл.: 4. Бібліогр.: 12 назв.

Ключові слова: визначення об'єктів; системи медичної відеозйомки; навчання та тестування; моделі; анотація зображень відеокадрів.

УДК 004.932.2

Совершенство этапов разработки модели определения объектов в системе медицинской видеосъемки / Белобородова Т.А., Приймак С.А., Коверга М.А. // Вестник НТУ "ХПИ". Серія: Информатика и моделирование. – Харьков: НТУ "ХПИ". – 2020. – № 1 (3). – С. 143 – 153.

Представлено решение задачи разработки модели определения объектов в системах видеосъемки медицинских процессов за счет установленной совокупности этапов. В ходе работы систематизирована совокупность этапов разработки модели, проведен эксперимент по предобработке, аннотации изображений видеокадров, обучения и тестирования модели на данных лапароскопической операции. Рассмотрены основные шаги обучения и тестирования модели с использованием сети Faster R-CNN. Точность на тестовых данных определена с использованием средней оценки точности модели и равна 96.3%. Ил.: 4. Библиогр.: 12 назв.

Ключевые слова: определение объектов; системы медицинской видеосъемки; обучение и тестирование; модели; аннотация изображений видеокадров.

UDC 004.932.2

A stages of the development of objects detection model for medical video system / Biloborodova T.O., Prymak S.O., Koverha M.O. // Herald of the National Technical University "KhPI". Series of "Informatics and Modeling". – Kharkov: NTU "KhPI". – 2020. – № 1 (3). – P. 143 – 153.

The solution of the problem of developing a model for objects detection in video systems of medical processes due to the established set of stages is presented. In the course of the work, the set of stages of model development was systematized, an experiment was conducted on preprocessing, annotation of video frame images, training and testing of the model using laparoscopic surgery data. The basic steps of training and testing the model using the Faster R-CNN network are considered. The accuracy on the test data was determined using an mean average precision of the model equal to 96.3%. Figs.: 4. Refs.: 12 titles.

Keywords: objects detection; medical video recording systems; training and testing; models; abstract images of video frames.