DOI: https://doi.org/10.33216/2222-3428-2019-17-3

УДК 004.932.72'1:622

Білобородова Т.О., Сіряк Р.В., Скарга-Бандурова І.С., Давіденко М.О., Сіроштан І.В., Приймак С.О.

РОЗПІЗНАВАННЯ ВЗІЄМОДІЇ ІНСТРУМЕНТІВ З ТКАНИНОЮ НА МЕДИЧНИХ ВІДЕОЗОБРАЖЕННЯХ

У статті досліджується технологія розпізнавання медичних зображень, що включає анотування відеозображень та їх використання для навчання моделі визначення об'єктів на відео. Розглянуто режими маркування відеозображень, технології розпізнавання з використанням методів визначення просторової активності об'єктів і сегментації. Представлений процес реалізації анотування медичних відеозображень лапароскопічного хірургічного втручання для виділення об'єктів розпізнавання. Досліджено процесс навчання моделі, структура і конфігурація мережі для створення моделі розпізнавання об'єктів медичних відеозображеннь. Представлений процес анотування медичних відеозображень лапароскопічного хірургічного втручання для виділення об'єктів розпізнавання. Визначені етапи підготовки та розпізнавання відеозображень: анотування відеозображень, навчання моделі, використання навченої моделі для нових нерозмічених відеозображень в режимі реального часу. Для навчання моделі використана мережа виявлення об'єктів в режимі реального часу YOLO. Компоненти виявлення об'єктів відеозображення об'єднуються в єдину нейронну мережу. Мережа використовує функції всього зображення для прогнозування кожної коробки. Класифікуються всі обмежуючі прямокутники коробки для всіх класів зображення одночасно. Навчання моделі проводилося з використанням відеозображень, анотованих для задач виявлення об'єктів та локалізації. Для оцінки якості моделі використаний параметр mean average precision (mAP). На 10000 ітерацій отримані наступні результати. Загальна кількість виявлень = 38154, з них правильних виявлень = 25248. Кількість хибно позитивних (FP), хибно негативних (FN), істинно позитивних (TP) і істинно негативних (TN) результатів виявлення розподілена наступним чином. TN = 7071, TP = 7656, FP = 5835, FN = 17592. Ці результати розпізнавання використані для розрахунку IoU і mAP. Середній показник IoU = 41,36%, mAP = 0.290665 або 29.07%..

***Ключові слова**: розпізнавання, відеозображення, анотування, сегментація, нейронна мережа.*

Актуальність дослідження. Розпізнавання взаємодії хірургічних інструментів з тканинами внутрішніх органів є важливим етапом аналізу лапароскопічних операцій. Такі операції здійснюються через природні отвори тіла або невеликі штучні розрізи, за рахунок чого зменшуються травми пацієнтів та скорочується час їх госпіталізації. Разом з тим, під час лапароскопічних операцій можливі обмеження бачення та рухливості, ускладнена координація рук і очей хірурга, що обумовлює створення нових засобів візуалізації, контролю та управління. З цього приводу, спільнота медичних дослідників прагне розширити можливості хірурга за допомогою контекстно-ведених комп'ютерних хірургічних систем. Метою таких систем є забезпечення хірурга актуальною інформацією, аналіз дій під час операції, прогнозування можливих непередбачуваних ситуацій з метою їх уникнення, підтримка прийняття рішення при діагностуванні патологічних аномалій внутрішніх органів, тощо. Одним із завдань, в контексті задачі аналізу дій хірурга під час операції, є аналіз відеозображень та розпізнавання взаємодії інструментів з тканинами. Якість диференціації об'єктів має вирішальне значення для оцінки продуктивності, оскільки існує багато факторів, що ускладнюють розпізнавання, серед яких роздільна здатність відеокамери, наявність у порожнині газів, диму, запотівання лінзи камери, індивідуальні особливості анатомії та інші. Також, обмежені сегменти реальних операцій можуть не дати всебічну оцінку хірургічного процесу.

Однією з передумов кількісної оцінки взаємодії тканина-інструмент, є сегментація хірургічних епізодів та виявлення деформації тканини у відповідь на рухи інструменту. Для отримання якісної, ефективної моделі необхідна попередня обробка відеозображень для отримання даних об'єктів. Такі дані можуть включати анотування присутності хірургічного інструменту або фази операції, які є мітками для більш детальних завдань, таких як семантична сегментація. У цій роботі ми пропонуємо підхід до сегментації та аналізу взаємодії хірургічних інструментів з тканинами, що засновано на декількох візуальних сигналах.

Аналіз останніх досліджень і публікацій. Типовий алгоритм розпізнавання об’єктів, як правило, поділяється на два етапи: отримання даних про об’єкти та класифікація об’єктів.

Метою отримання даних об’єктів є перетворення відеозображення у векторний вигляд, витяг репрезентативної та дискримінативної інформації про об’єкти та мінімізація можливих варіацій. Отримання даних про об’єкти є першою і найважливішою задачею розпізнавання об’єктів. Об’єкти, що з’являються у відео, відрізняються за своєю швидкістю просторової активності, кутом огляду камери, зовнішнім виглядом та варіаціями сцени. Якісні методи отримання даних про об’єкти повинні бути ефективними для обчислення, ефективними для характеристики об’єктів та повинні максимально збільшувати розбіжність між об’єктами для зменшення помилки класифікації [1]. Отримання даних про об’єкт може включати [2]:

- виявлення і відстеження просторової активності, тобто руху об’єктів – дані координат об’єктів;

- виявлення властивостей об’єктів – дані сегментації об’єктів.

Дані просторової активності і сегментації об’єктів зазвичай відстежуються від одного кадру до іншого в послідовності зображень. Існують також методи, що направлені на виявленні контурів об’єктів у відео [3], але їх використання для розпізнавання медичних відеозображень не є доцільним з причини важливості властивостей даних сегментації об’єкта при розпізнаванні діагностичних або анатомічних особливостей у відеозображеннях. Підходи отримання даних об'єктів відео для навчання моделі поділяють [4] на повністю автоматичні методи, при використанні яких на вхід мережі подається тільки відео, і напівавтоматичні методи, які вимагають попереднього початкового аннотування відео об'єктів. При використанні напівавтоматичного підходу використовується попереднє анотування відео об'єктів, тобто співставлення вхідному вектору даних об’єкта вихідної мітки – класу об’єкта Користувач повинен спочатку визначити координати, дані сегментації об'єкта, за допомогою яких наступні кадри можуть бути автоматично анотовані і сегментовані.

Анотування відеозображення - це процес маркування зображень ідентифікаторами (мітками). Чим складніше завдання, тим точніше анотовані дані потрібні для високоякісного навчання. Крім того, навчання має враховувати зворотний зв'язок в реальному часі, щоб зменшити кількість помилок і розширити можливості алгоритму визначення об'єктів. Анотування відеозображень для визначення об'єктів допомагає реалізувати наступні можливості:

- визначення об’єктів в зображенні;
- визначення місцезнаходження кожного об’єкту в певній сцені;
- порівняння кольору, розміру і форми об’єктів;
- розпізнавання точок подібності між двома та більше зображеннями;
- розпізнавання різниці між двома або більше зображеннями;
- розуміння безперервності в зображеннях, якщо вона є.

Під час застосування навченої моделі після обчислення даних об’єктів модель обирає необхідну мітку для об’єкта. Моделі об’єктів можна розділити за наступними категоріями:

Пряме прогнозування. У цьому випадку модель підсумовує наданні вектори, а потім розпізнає об’єкти за допомогою стандартних алгоритмів класифікації. У цих методах динаміка просторової активності об’єкту характеризується цілісно, використовуючи форму просторової активності або кодує розподіл локальних моделей просторової активності за допомогою гістограми.

Прогнозування послідовностей. При використанні цих підходів прогнозується часова зміна просторової активності або створюється за допомогою моделей послідовностей, таких як приховані Марковські моделі, умовні випадкові поля (CRF), структурований алгоритм опорних векторів (SSVM) [5]. Ці підходи сприймають відео як композицію часових фрагментів або кадрів. Модель аналізує траєкторію просторової активності об’єкта для класифікації. Цей метод відкидає ряд часових неінформативних позицій за тимчасовим наслідком і створює більш компактну послідовність для класифікації. Тим не менш, ці послідовні підходи переважно використовують цілісний характер.

Змішане прогнозування. Цей метод більш актуальний для структурованих об’єктів. Наприклад, якщо розглядати хірургічну порожнину як структурований об’єкт то можна легко моделювати рухи об’єктів. Частково-базові підходи розглядають інформацію про просторову активність як з усього об'єкта так і з його частин [2]. Перевага цієї категорії підходів полягає в тому, що вона по суті фіксує геометричні відносини між частинами об’єкта, що є важливою підказкою для розрізнення їх просторової активності.

Нещодавні дослідження [6] показали, що особливості просторової активності об’єктів можуть бути досліджені за допомогою методів глибокого навчання, таких як згорткові нейронні мережі (CNN) та рекурентні нейронні мережі (RNN). Використовуючи RGB-кадри та оптичні потоки кадрів, нейронні мережі показали якісні результати на різних наборах даних про дії. Автори дослідження [7] представили підхід, заснований на згорткових нейронних мережах, для вирішення завдання просторового виявлення інструментів в реальних лапароскопічних хірургічних відеороликах. Представлений підхід показав високу продуктивність просторового виявлення об'єктів в режимі реального часу. В роботі [8] автори дослідили метод сегментації і розпізнавання хірургічних інструментів в відеозаписах, лапароскопічних гінекологічних втручань. Авторами оцінена досяжна продуктивність сегментування хірургічних інструментів по масці їх фону. Використана згорткова мережа для бінарної сегментації та розпізнавання хірургічних інструментів та мультикласова сегментація для розпізнавання типу інструменту. По результатам дослідження на невеликій кількості навчальних прикладів отримана висока точність бінарного розпізнавання, тобто точність розпізнавання інструментів. Але при мультикласовому розпізнаванні визначити тип інструмент все ще дуже складно через високу схожість хірургічних інструментів між собою. Слід зазначити, що більшість методів розпізнавання ефективні при використанні для передчасно зроблених відеозаписів. Їх продуктивність при використанні в умовах неповних даних має значно гірші результати.

Беручі до уваги вищевказане, для розпізнавання динамічних медичних відеозображень в режимі реального часу пропонується використання змішаних підходів, заснованих на використанні даних координат об’єктів у послідовностях відеозображень та даних сегментації об’єктів, узагальнених для навчання моделі розпізнавання.

Метою статті є дослідження технології розпізнавання відеозображень та їх використання для оцінки взаємодії тканина-інструмент. Для досягнення мети поставлені наступні завдання.

1. Аналіз технологій обробки медичних відеозображень.

2. Визначення загальної структури процесу розпізнавання медичних зображень.

3. Дослідження процесу навчання моделі розпізнавання медичних відеозображень з використанням нейронної мережі та визначення її конфігурації.

4. Анотування медичних відеозображень лапароскопічного хірургічного втручання для виділення хірургічніх інструментів та внутрішніх органів під час операції, отримання даних координат та сегментації об’єктів, що описують хірургічні інструменти та внутрішні органи, які присутні в кожному відеокадрі;

5. Навчання моделі з використанням даних координат і сегментації об'єктів медичних відеозображень та оцінка якості моделі.

Метод

Загальна структура процесу розпізнавання медичних зображень

Підготовка і розпізнавання відеозображень містить три наступні етапи, як це представлено на рис. 1.

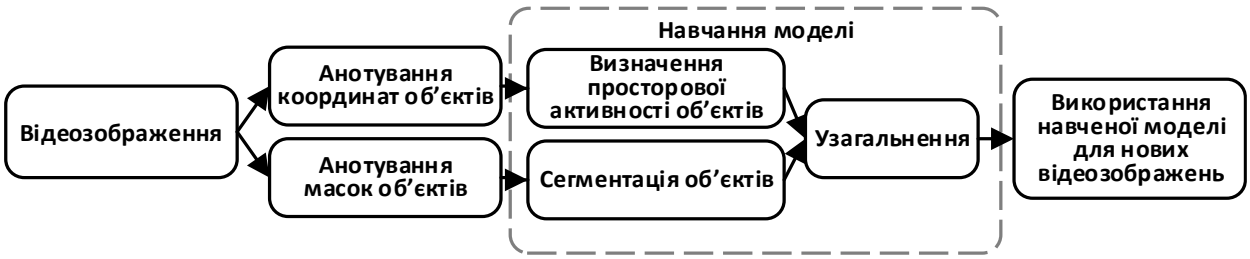


Рисунок 1 - Етапи підготовки і розпізнавання відеозображень

1. Анотування відеозображень, що містить визначення координат маркованих об'єктів на кожному відеокадрі та виділення об'єктів на кожному відеокадрі.

2. Навчання моделі, що передбачає визначення просторової активності об’єктів з використанням анотованих координат обмежуючого прямокутника, відповідних кожному класу, сегментацію з використанням даних масок анотованих об'єктів, цілочисельне значення для кожного зображення та узагальнення – об’єднання узагальненого уявлення об'єктів.

3. Використання навченої моделі для нових нерозмічених відеозображень в режимі реального часу.

Процесс навчання моделі для розпізнавання відеозображень

Компоненти виявлення об'єктів відеозображення об'єднуються в єдину нейронну мережу [9]. Мережа ділить зображення на своєрідну сітку і прогнозує обмежуючий прямокутник (bounding box) і ймовірності того, що там є шуканий об'єкт для кожної ділянки. Мережа використовує функції всього зображення для прогнозування кожного прямокутника. Прогнозуються всі обмежуючі прямокутники для всіх класів для зображення одночасно. Система розпізнавання ділить вхідне зображення на сітку S × S. Якщо центр об’єкта потрапляє в ділянку сітки, ця ділянка сітки відповідає за виявлення цього об'єкта. Кожна ділянка сітки прогнозує обмежуючий прямокутник *B* і достовірність прогнозування *Conf*. Ці оцінки достовірності відображають те, наскільки модель впевнена в тому, що прямокутник містить об'єкт, а також те, наскільки точним він вважає прогнозований прямокутник. Достовірність *Conf* можна визначити наступним чином.

$$Conf = \Pr(Object) * \mathrm{IoU}_{pred}^{truth},$$

де $\Pr(Object)$ - ймовірність того, що прямокутник містить об'єкт; IoU - узагальнення достовірності (IoU) між прогнозованим і справжнім прямокутниками. IoU (Intersection over Union) – метріка, що відображає відношення площі перетину коробки, отриманої в результаті детектування та коробки з анотування до площі їх об'єднання.

$$\mathrm{IoU}_{pred}^{truth} = \frac{pred \cap truth}{pred \cup truth}.$$

Якщо в ділянці немає об'єктів, оцінки довіри повинні дорівнюватися нулю. В іншому випадку, показник достовірності дорівнює узагальненню достовірності (IoU) між прогнозованим і справжнім прямокутником.

Кожен обмежуючий прямокутник *B* містить п'ять прогнозованих змінних: *x, y, w, h* та Pr прямокутнику. Координати (*x, y*) представляють центр прямокутника *B* щодо границь ділянки сітки. Прямокутник обмежується шириною *w* і висотою *h* прогнозованих щодо всього зображення. Показник ймовірності Pr відображає ступінь ймовірності того, що прямокутник містить об'єкт і наскільки точним є обмежуючий прямокутник. Кожна ділянка сітки також прогнозує ймовірності умовного класу *C*, $\Pr(Class_i|Object)$. Ці ймовірності обумовлені ділянкою сітки, що містить об’єкт. Для кожної ділянки сітки прогнозується тільки один набір ймовірностей класів, незалежно від кількості прямокутників *B*.

При проведенні тестового прогнозування умовні ймовірності класу множаться на прогнозовану достовірність прямокутника:

$$\Pr(Class_i \mid Object) * \Pr(Object) * \mathrm{IoU}_{pred}^{truth} = \Pr(Class_i) * \mathrm{IoU}_{pred}^{truth}, \quad (1)$$

де $\Pr(Class_i \mid Object)$ - ймовірність того, що об'єкт належить класу *i*, якщо об'єкт присутній в прямокутнику; $\Pr(Class_i)$ - ймовірність того, що об'єкт належить класу *i*. Цей вислів допомагає визначити достовірність класифікації та локалізації об'єкта.

Останній шар мережі прогнозує ймовірності класу, і координати обмежуючого прямокутника. Проводиться нормалізація ширини і висоти обмежуючого прямокутника по ширині і висоті зображення до шкали від 0 до 1. Координати обмежуючого прямокутника *x* і *y* параметризуються для подання їх від 0 до 1.

Для кінцевого шару використовується лінійна функція активації, а всі інші шари використовують наступну лінійну активацію.

$$\phi(x) = \begin{cases} x, \text{ якщо } x > 0 \\ 0.1x, \text{ в іншому випадку} \end{cases}. \quad (2)$$

Параметром якості отриманої моделі визначена сума квадратів помилок. Крім того в кожному зображенні багато ділянок сітки, що не містять ніяких об'єктів. Це наближає значення достовірності цих ділянок до 0. Це може привести до нестабільності моделі, що призведе до того, що навчання на ранніх стадіях буде невідповідним. Щоб уникнути нестабільності, можливе збільшення втрат прогнозувань координат обмежуючого прямокутника і зменшення втрат прогнозувань достовірності для прямокутників, що не містять об'єктів. Для цього використовуються два параметри, λ_{coord} і λ_{noobj}.

Сума квадратів помилок також пропорційна розміру прямокутника. Невеликі відхилення у великих ділянках мають менше значення, ніж в маленьких прямокутниках. Для вирішення цієї проблеми використовується прогнозований корінь квадратний з ширини і висоти обмежуючого прямокутника.

Значення похибки обчислюється функцією втрат наступним чином.

$$\begin{aligned}
&\lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{N}_{ij}^{obj} \left[(x_i - x_i)^2 + (y_i - y_i)^2 \right] + \\
&+\lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{N}_{ij}^{obj} \left[(\sqrt{w_i} - \sqrt{w_i})^2 + (\sqrt{h_i} - \sqrt{h_i})^2 \right] + \quad , \\
&+\sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{N}_{ij}^{obj} (C_i - C_i)^2 + \lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{N}_{ij}^{noobj} (C_i - C_i)^2 + \\
&+\sum_{i=0}^{S^2} \mathbb{N}_{i}^{obj} \sum_{c \in \text{classes}}^{B} (p_i(c) - p_i(c))^2
\end{aligned}$$

де $\mathbb{N}_i^{obj}$ позначає, якщо об'єкт з'являється в ділянці *i*, а $\mathbb{N}_{ij}^{obj}$ позначає, що *j*-й предікатор обмежуючого прямокутника в ділянці *i* відповідає за це прогнозування. C_i - довірча ймовірність прямокутника *j* в ділянці *i*. $\mathbb{N}_i^{obj} = 1$ означає, що об'єкт знаходиться в ділянці *i*, в іншому випадки 0. $\mathbb{N}_{ij}^{noobj}$ є доповненням $\mathbb{N}_{ij}^{obj}$. $p_i(c)$ - умовна ймовірність класу *c* в ділянці *i*. $\mathbb{N}_{ij}^{obj} = 1$ якщо *j*-тий обмежуючий прямокутник ділянки *i* відповідає за виявлення об'єкта, в іншому випадку 0. λ_{coord} - збільшує вагу за втрату в координатах обмежуючого прямокутника. λ_{noobj} - знижує вагу при виявленні фону.

Функція втрат штрафує за помилку класифікації, якщо об'єкт присутній в цій ділянці сітки. Вона також штрафує тільки помилку координат прямокутника, якщо цей предиктор відповідає за базовий прямокутник істинності (тобто має найвищий IoU з усіх предикторов в цій ділянці сітки).

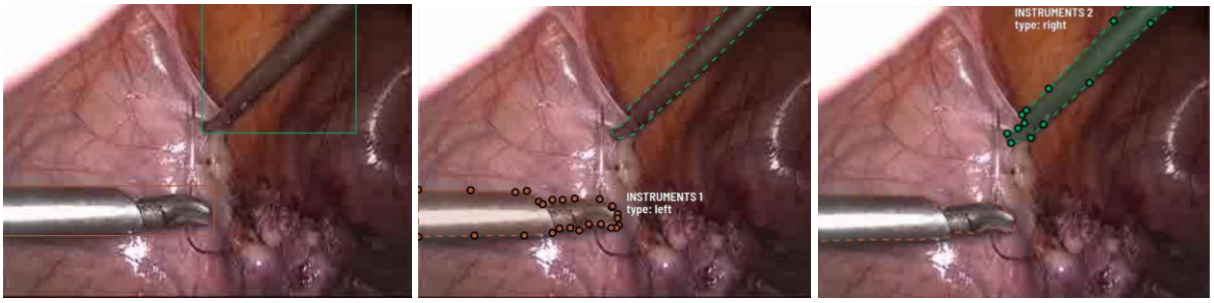
Конфігурація мережі

Налаштування параметрів мережі проведено з використанням наступних параметрів конфігурації. Batch - кількість зразків, які будуть попередньо оброблені в одній партії. Параметр subdivisions показує кількість ділянок сітки, на яке розбивається batch для паралельної обробки. Параметр decay є навчальним параметром і використовується для стабільності роботи мережі і контролює зменшення ваги для уникнення великих значень. Параметр momentum є навчальним параметром і робить градієнт більш стабільним. Новий градієнт обчислюється наступним чином: momentum*попередній_градієнт+(1-momentum)*градієнт_поточного_batch. Параметр channel визначає розмір мережі (канали). Кожне зображення буде перетворено в визначену кількість каналів під час навчання і виявлення.

Результати

Для анотування об’єктів відеозображень було застосовано пакет програмного забезпечення CVAT [10]. Доступні режими анотації відеозображень дозволяють анотування координат обмежуючих прямокутників і масок об'єктів.

Виконано анотування об'єктів медичних відеозображень. Приклад процесу анотування медичних відеозображень з визначенням координат хірургічних інструментів, що використовуються при проведенні лапароскопічної операції, і маски об'єктів з використанням обмежуючого прямокутника та полігональної рамки об'єкта представлений на рис. 2.



а) б) в)

Рисунок 2 - Приклад процесу анотування об'єктів медичних відеозображень з використанням а) обмежуючого прямокутника та б), в) полігональної рамки

Навчання моделі виконано на відеозображеннях, анотованих з використанням обмежуючого прямокутника для задач виявлення об'єктів і локалізації. Анотовані дані складаються з 8 класів об'єктів, Початковий розмір відеозображень 512 × 512. Для навчання моделі використана мережа для виявлення об'єктів в режимі реального часу YOLO [11]. Використана наступна конфігурація мережі для навчання моделі: batch = 64, subdivisions = 64, channels = 3, momentum = 0.9, decay = 0.0005.

Для оцінки якості моделі використаний параметр mean average precision (mAP). Середня середня точність для набору запитів - це середнє значення середніх показників точності для кожного запиту.

$$MAP = \frac{\sum_{q=1}^{Q} AveP(q)}{Q},$$

де Q - кількість запитів в наборі, а $AveP\ (q)$ - середня точність (Average Precision) для даного запиту q.

На 10000 ітерацій отримані наступні результати. Загальна кількість виявлень = 38154, з них правильних виявлень = 25248. Кількість ложно позитивних (FP), помилково негативних (FN), істинно позитивних (TP) і істинно негативних (TN) результатів виявлення представлене в такий спосіб. TN = 7071, TP = 7656, FP = 5835, FN = 17592. Ці результати розпізнавання використані для розрахунку IoU і mAP. Середній показник IoU = 41,36%, mAP = 0.290665 або 29.07%.

Висновки. Досліджено технологію розпізнавання медичних відеозображень з проведенням анотації відеозображень, використання анотованих відеозображень для створення моделі визначення об’єктів на відео. Розглянуто режими анотування відеозображень, технології розпізнавання з використанням методів визначення просторової активності і сегментації об’єктів. Представлений процес реалізація анотування медичних відеозображень лапороскопічного хірургічного втручання для визначення об’єктів розпізнавання. Досліджено процес навчання моделі, структура і конфігурація мережі для створення моделі розпізнавання об'єктів на медичних відеозображеннях.

Навчання моделі проведено з використанням відеозображень, анотованих для задач виявлення об'єктів і локалізації. Для оцінки якості моделі використаний параметр mean average precision (mAP). На 10000 ітерацій загальна кількість виявлень склало 38154, з них правильних виявлень = 25248. На 10000 ітерацій mAP = 0.290665 або 29.07%.

Література

1. Jo, K., Choi, Y., Choi, J. and Chung, J.W., 2019. Robust Real-Time Detection of Laparoscopic Instruments in Robot Surgery Using Convolutional Neural Networks with Motion Vector Prediction. Applied Sciences, 9(14), p.2865.
2. Mishra, M.S.K., Jtmcoe, F. and Bhagat, K.S., 2015. A survey on human motion detection and surveillance. International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE) Volume, 4.
3. Xie, S. and Tu, Z., 2015. Holistically-nested edge detection. In Proceedings of the IEEE international conference on computer vision. P. 1395-1403.
4. Chen, Y., Hao, C., Wu, W., & Wu, E. (2017). Efficient frame-sequential label propagation for video object segmentation. Multimedia Tools and Applications, 77 (5), 6117-6133. doi: 10.1007 / s11042-017-4520-5
5. F. Lalys, L. Riffaud, D. Bouget, and P. Jannin. A framework for the recognition of high-level surgical tasks from video images for cataract surgeries. IEEE Transactions on Biomedical Engineering, 59(4):966–976, 2012
6. Kong, Y. and Fu, Y., 2018. Human action recognition and prediction: A survey. arXiv preprint arXiv:1806.11230.

7. Jin, A., Yeung, S., Jopling, J., Krause, J., Azagury, D., Milstein, A. and Fei-Fei, L., 2018, March. Tool detection and operative skill assessment in surgical videos using region-based convolutional neural networks. In 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). P. 691-699. IEEE.
8. Kletz, S., Schoeffmann, K., Benois-Pineau, J. and Husslein, H., 2019, September. Identifying Surgical Instruments in Laparoscopy Using Deep Learning Instance Segmentation. In 2019 International Conference on Content-Based Multimedia Indexing (CBMI). P. 1-6. IEEE.
9. Redmon, J., Divvala, S., Girshick, R. and Farhadi, A., 2016. You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition. P. 779-788.
10. Computer Vision Annotation Tool (CVAT). URL: https://github.com/opencv/cvat (дата звернення: 15.11.2019)
11. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi: 10.1109 / cvpr.2016.91

References

1. Jo, K., Choi, Y., Choi, J. and Chung, J.W., 2019. Robust Real-Time Detection of Laparoscopic Instruments in Robot Surgery Using Convolutional Neural Networks with Motion Vector Prediction. Applied Sciences, 9(14), p.2865.
2. Mishra, M.S.K., Jtmcoe, F. and Bhagat, K.S., 2015. A survey on human motion detection and surveillance. International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE) Volume, 4.
3. Xie, S. and Tu, Z., 2015. Holistically-nested edge detection. In Proceedings of the IEEE international conference on computer vision. P. 1395-1403.
4. Chen, Y., Hao, C., Wu, W., & Wu, E. (2017). Efficient frame-sequential label propagation for video object segmentation. Multimedia Tools and Applications, 77 (5), 6117-6133. doi: 10.1007 / s11042-017-4520-5
5. F. Lalys, L. Riffaud, D. Bouget, and P. Jannin. A framework for the recognition of high-level surgical tasks from video images for cataract surgeries. IEEE Transactions on Biomedical Engineering, 59(4):966–976, 2012
6. Kong, Y. and Fu, Y., 2018. Human action recognition and prediction: A survey. arXiv preprint arXiv:1806.11230.
7. Jin, A., Yeung, S., Jopling, J., Krause, J., Azagury, D., Milstein, A. and Fei-Fei, L., 2018, March. Tool detection and operative skill assessment in surgical videos using region-based convolutional neural networks. In 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). P. 691-699. IEEE.
8. Kletz, S., Schoeffmann, K., Benois-Pineau, J. and Husslein, H., 2019, September. Identifying Surgical Instruments in Laparoscopy Using Deep Learning Instance Segmentation. In 2019 International Conference on Content-Based Multimedia Indexing (CBMI). P. 1-6. IEEE.
9. Redmon, J., Divvala, S., Girshick, R. and Farhadi, A., 2016. You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition. P. 779-788.
10. Computer Vision Annotation Tool (CVAT). URL: https://github.com/opencv/cvat (дата звернення: 15.11.2019)
11. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi: 10.1109 / cvpr.2016.91.

В статье исследуется технология распознавания медицинских видеоизображений, включая аннотацию видеоизображений и использования аннотированных видеоизображений для создания модели определения пространственной активности и сегментации объектов в видео. Рассмотрены режимы аннотирования видеоизображений, технологии распознавания с использованием методов обнаружения объектов и сегментации. Представлен процесс реализации аннотации медицинских видеоизображений лапороскопичного хирургического вмешательства для выделения объектов распознавания. Представлены процесс обучения модели, структура и конфигурация сети для создания модели распознавания объектов медицинских видеоизображений. Представлен процесс аннотирования медицинских видеоизображений лапароскопического хирургического вмешательства для определения объектов распознавания. Определены этапы подготовки и распознавания видеоизображений: аннотирования видеоизображений, обучения модели, использование обученной модели для новых неразмеченную видеоизображений в режиме реального времени. Для обучения модели использована сеть обнаружения объектов в режиме реального времени YOLO. Компоненты обнаружения объектов видеоизображения объединяются в единую нейронную сеть. Сеть использует функции всего изображения для прогнозирования каждого прямоугольника. Классифицируются все ограничивающие прямоугольники для всех классов изображения одновременно. Обучение модели проведено с использованием видеоизображений, аннотированных для задач обнаружения объектов и локализации. Для оценки качества модели использован параметр mean average precision (mAP). На 10000 итераций получены следующие результаты. Общее количество обнаружений = 38154, из них правильных выражений = 25248. Количество ложно положительных (FP), ложно отрицательных (FN), истинно положительных (TP) и истинно отрицательных (TN) результатов выявления распределена следующим образом. TN = 7071, TP = 7656, FP = 5835, FN = 17592. Эти результаты распознавания использованы для расчета IoU и mAP. Средний показатель IoU = 41,36%, mAP = 0.290665 или 29.07%.

Ключевые слова: *распознавание, видеоизображения, аннотирование, сегментация, нейронная сеть*

The technology for recognizing medical video images is described. The video annotation and using of annotated video to model creation for identifying objects in video are present. The modes of video labeling, recognition technologies using methods of object detection and segmentation are considered. The process of implementation of the annotation of medical video images of laparoscopic surgical intervention to highlight recognition objects is presented. The mathematical approach, the structure and configuration of the network are studied to create a model for recognizing objects of medical video. The annotation process of laparoscopic surgery video for objects recognition is presented. The stages of preparation and recognition of video images are determined: annotation of video images, model training, use of the trained model for new unlabeled video images in real time. For model training, the YOLO real-time object detection system was used. Components for detecting video objects are integrated into a single neural network. The network uses the functions of the entire image to predict each rectangle. All bounding box for all image classes are classified simultaneously. The model was trained using video images annotated for objects detection and localization. To assess the quality of the model, the mean average precision (mAP) parameter was used. At 10,000 iterations, the following results were obtained. The total number of detections = 38154, of which the correct expressions = 25248. The number of false positive (FP), false negative (FN), true positive (TP) and true negative (TN) detection results is distributed as follows. TN = 7071, TP = 7656, FP = 5835, FN = 17592. These recognition results were used to calculate IoU and mAP. Average IoU = 41.36%, mAP = 0.290665 or 29.07%.
***Keywords**: video, recognition, annotation, semantic segmentation, neural network.*

Т.О. Білобородова – доцент кафедри комп'ютерних наук та інженерій СНУ ім. В.Даля, кандидат технічних наук, ORCID ID: 0000-0001-7561-7484
Р.В. Сіряк – пошукач кафедри комп'ютерних наук та інженерій СНУ ім. В.Даля
І.С. Скарга-Бандурова – завідувач кафедри комп'ютерних наук та інженерій СНУ ім. В.Даля, професор, доктор технічних наук, ORCID ID: 0000-0003-3458-8730
М.О. Давіденко – магістр кафедри комп'ютерних наук та інженерій СНУ ім. В.Даля
І.В. Сіроштан – магістр кафедри комп'ютерних наук та інженерій СНУ ім. В.Даля
С.О. Приймак – магістр кафедри комп'ютерних наук та інженерій СНУ ім. В.Даля