DOI 10.31891/2307-5732-2023-329-6-247-252
УДК 004.8

МУЛЯР ЕДУАРД
Хмельницький національний університет
https://orcid.org/0000-0003-4052-4696
e-mail: edikmulyar228@gmail.com
БАГРІЙ РУСЛАН
Хмельницький національний університет
https://orcid.org/0000-0001-5219-1185
e-mail: bahriiro@khmnu.edu.ua
ПАСІЧНИК ОЛЕКСАНДР
Хмельницький національний університет
https://orcid.org/0000-0002-8760-4688
e-mail: o.a.pasichnyk@gmail.com
МАНЗЮК ЕДУАРД
Хмельницький національний університет
https://orcid.org/0000-0002-7310-2126
e-mail: eduard.em.km@gmail.com

МЕТОД ВИЯВЛЕННЯ ЗОВНІШНІХ ПРОЯВІВ НАСИЛЬСТВА У ВІДЕОПОТОЦІ НЕЙРОМЕРЕЖЕВИМИ ЗАСОБАМИ

Проблема виявлення проявів насильства за зображеннями у відеопотоці є актуальною в сучасному світі зі зростаючою кількістю відеоматеріалів, що містять насильницькі сцени. Це включає відео, зняте на вулицях, в громадських місцях та відеозаписи з камер спостереження. Виявлення та реагування на такі сцени важливі для забезпечення безпеки у громадських просторах та захисту прав людини.

Для інтелектуалізації процесу відеоспостереження сьогодні активно використовуються інформаційні технології, а саме нейромережі. Застосування нейромережевих засобів у відеоспостереженні є важливим засобом, оскільки дозволяє автоматично аналізувати великі обсяги відеоматеріалів і виявляти насильницькі сцени з високою точністю.

У статті пропонується метод виявлення зовнішніх проявів насильства за зображеннями у відеопотоці за допомогою згорткової нейронної мережі та класифікатора SVM. На вхід методу подаються кадри відеоматеріалу з яких згорткова нейронна мережа вилучає набір ознак, який потім передається класифікатору SVM для отримання оцінки щодо ймовірності належності цих ознак до певного класу (насильницького або не насильницького). Особливістю запропонованого методу є можливість працювати із відеоматеріалом у реальному часі. Це досягається за рахунок того, що згорткова нейронна мережа використовуючи метод fine-tuning навчалася на неперервному потоці даних із мультимедійних платформ для онлайн трансляцій.

Проведено експерименти з використанням різних наборів даних для оцінки ефективності запропонованого методу. Результати показали, що метод досягає високої точності (87,4%-99,45%) виявлення насильства та працює ефективно з відеопотоком даних у реальному часі.

Ключові слова: насильство, виявлення, відеопотік, нейромережі, згорткова нейронна мережа, SVM.

EDUARD MULIAR, RUSLAN BAHRII, ALEXANDER PASICHNUK, EDUARD MANZIUK
Khmelnytskyi National University

METHOD OF DETECTING OUTWARD MANIFESTABONS OF VIOLENCE IN VIDEO STREAMS USING NEURAL NETWORK TOOLS

The problem of detecting violence from images in a video stream is relevant in today's world with a growing number of videos containing violent scenes. This includes video taken on the streets, in public places, and from surveillance cameras. Identifying and responding to such scenes is important for ensuring safety in public spaces and protecting human rights. Information technologies, namely neural networks, are being actively used to intellectualize the video surveillance process. The use of neural network tools in video surveillance is an important tool, as it allows to automatically analyze large amounts of video materials and detect violent scenes with high accuracy.

The article proposes a method for detecting external manifestations of violence in images in a video stream using a convolutional neural network and an SVM classifier. The input to the method is video frames from which the convolutional neural network extracts a set of features, which is then passed to the SVM classifier to obtain an estimate of the probability of these features belonging to a certain class (violent or non-violent). The peculiarity of the proposed method is the ability to work with video material in real time. This is achieved due to the fact that the convolutional neural network was trained using the fine-tuning method on a continuous stream of data from multimedia platforms for online broadcasts. Experiments were conducted using different datasets to evaluate the effectiveness of the proposed method. The results showed that the method achieves high accuracy (87,4%-99,45%) in detecting violence and works efficiently with a real-time video data stream.

The use of neural network tools to detect violence in a video stream has great potential in various fields, including public safety, cybersecurity, and human rights protection. Improving the proposed method can help to expand the possibilities of detecting and preventing violence in video streams.

Keywords: violence, detection, video stream, neural networks, convolutional neural network, SVM.

Постановка проблеми

Сьогодні для протидії такій суспільній проблемі як насильство розпочали активно використовувати системи відеоспостережень. Такі країни, як Китай та Південна Корея, є лідерами у встановленні камер

відеоспостереження, і результати використання цих систем є вражаючими. У Китаї рівень насильства у громадських місцях знизився на 60%, а в Південній Кореї – на, приблизно, 50% [1].

Однак, ці системи мають деякі недоліки. Основною проблемою є людський фактор, зокрема неуважність та недбалість операторів-спостерігачів. Зазвичай оператор може ефективно контролювати лише обмежену кількість камер відеоспостереження. Однак, коли кількість камер перевищує межі сприйняття, оператори можуть допускати помилки або пропускати випадки насильства.

Для вирішення цієї проблеми сьогодні активно використовуються передові інформаційні технології, зокрема штучні нейронні мережі. Завдяки цим технологіям, системи відеоспостереження можуть автоматично аналізувати великий обсяг відеоданих і виявляти потенційні випадки насильства. Штучний інтелект допомагає зменшити навантаження на операторів і забезпечує більш точне та ефективне виявлення зазначених подій [2].

Застосування штучного інтелекту в системах відеоспостереження є важливим кроком у забезпеченні безпеки у громадських місцях та протидії насильству. Ці технології допомагають забезпечити швидке реагування на зазначені події та вчасне виявлення потенційних загроз. Використання штучного інтелекту в системах відеоспостереження покращує загальний рівень безпеки та сприяє створенню безпечніших громадських просторів.

Аналіз останніх джерел

Для реалізації інтелектуального відеоспостереження сьогодні активно використовують такі моделі нейронних мереж як згорткова та рекурентна [3]. Згорткова нейронна мережа (CNN) є типом нейронних мереж, які широко використовуються для обробки зображень та роботи з багатомірними даними. Вони були розроблені саме для розпізнавання зображень та виконання завдань комп'ютерного зору [4]. Рекурентна нейронна мережа (RNN) є типом штучних нейронних мереж, які використовуються для моделювання послідовних даних, таких як мовний текст, часові ряди або музика, а також розпізнавання залежностей та шаблонів у цих даних. Основна відмінність RNN від інших типів нейронних мереж полягає в тому, що вона здатна зберігати попередній контекст та використовувати його для обробки наступних вхідних даних [5].

Зазвичай, наведенні моделі нейромереж не використовуються у «чистому» вигляді, як правило вони виступають центральним ядром яке модифікують або доповнюють іншими методами та моделями. Прикладом такого підходу є модель нейронних мереж BiConvLSTM (Bidirectional Convolutional LSTM). Ця модель поєднує в собі два потужних компоненти: двосторонню згорткову мережу (BiConv) та рекурентну нейронну мережу LSTM (Long Short-Term Memory). BiConvLSTM використовується для аналізу послідовних даних, таких як зображення або відео, з метою виявлення шаблонів та залежностей у цих даних. Основна ідея BiConvLSTM полягає в поєднанні двосторонньої згорткової мережі та LSTM для аналізу просторової та часової інформації в послідовних даних. Двостороння згорткова мережа використовується для виділення локальних ознак з різних частин зображення або відео. Вона дозволяє аналізувати дані як вперед (зліва направо) так і назад (справа наліво), тобто нейромережа може бачити як попередні пікселі в зображенні, так і наступні пікселі, що дозволяє мережі краще розуміти, як різні частини зображення пов'язані між собою. Відповідно це дозволяє аналізувати контекст з обох сторін та виявляти шаблони, які можуть бути присутніми в різних частинах послідовних даних [6].

Приклад використання моделі BiConvLSTM наведено у роботі [7], де розглянуто підхід до виявлення насильства у відео за допомогою методу «просторово-часовий кодер». «Просторово-часовий кодер» є методом який побудований за кількома архітектурами (BiConvLSTM, VGG13 [8]) для кодування кожного відеокадру як набору карт функцій. Ці карти функцій потім передаються до BiConvLSTM для подальшого кодування у часовому напрямку відео. Після цього виконується поелементна максимізація кожного з цих кодувань, щоб створити подання всього відео. Це подання передається класифікатору, щоб визначити, чи містить відео насильство. Щодо результатів роботи даного підходу, то для набору даних «Hockey fights» точність склала 96.54%, для набору даних «Violent flows» - 92.18%.

Іншим прикладом використання нейронних мереж для задачі виявлення насильства у відеопотоці є 3D CNN (3D Convolutional Neural Network). Дана модель є згортковою нейронною мережею, що використовується для обробки тримірних даних, таких як відео, медичні зображення або тримірні моделі. Основна ідея 3D CNN полягає у використанні тримірних згорток для виявлення просторових особливостей даних. Вона подібна до звичайних 2D CNN, що використовуються для обробки зображень, але має додатковий третій вимір для роботи з даними [9].

У роботі [10] запропоновано метод до виявлення насильства у відео за допомогою 3D-CNN. 3D-CNN спочатку обробляє кожний кадр відео, використовуючи набори фільтрів для виявлення важливих ознак, таких як рух, форма та колір. Потім 3D-CNN обробляє послідовність кадрів, використовуючи 3D-фільтри. Це дозволяє 3D-CNN виявляти динамічні ознаки насильства, такі як рухи тіла та взаємодії між людьми. Бінарний класифікатор використовується для класифікації насильства або його відсутності. Класифікація здійснюється шляхом застосування логістичної регресії до вихідного тензора 3D-CNN. На рахунок результатів роботи даного методу, то для набору даних «Hockey fights» точність склала 98.3%, для набору даних «Violent flows» - 97.17%.

В наведених методах виявлення насильства у відео не проведено дослідження з відеоматеріалами в реальному часі, що обмежує їх застосування в реальних задачах: автоматичного сповіщення про насильство

або покращення роботи операторів спостереження.

Метою роботи є розробка методу для виявлення зовнішніх проявів насильства у відеопотоці за допомогою нейромережевих засобів у реальному часі. Метод повинен працювати як зі статичним відеоматеріалом (відеоролик) так і з динамічним (відеопотік в реальному часі). Потрібно здійснити аналіз ефективності роботи запропонованого методу на відповідних наборах даних.

Виклад основного матеріалу

Робота запропонованого методу полягає в отриманні ознак насильства з кадрів вхідного відео за допомогою згорткової нейронної мережі і визначення ступеню насильства у відсотковому відношенні у відеопотоці за допомогою SVM (методу опорних векторів). На рис. 1 зображено архітектуру даного методу.

Етап 1 – Отримання кадрів із вхідного відеоматеріалу

Необхідно розбити вхідний відеоматеріал на послідовність кадрів та перетворити кожен кадр у карту зображень.

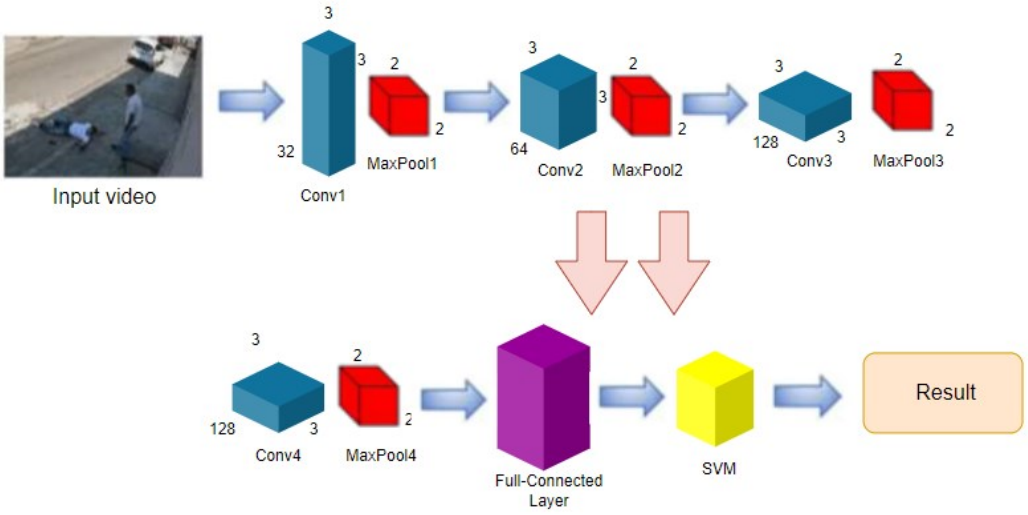

Рис. 1. Архітектура запропонованого методу

Етап 2 – Операція згортки

На даному етап необхідно виконати операцію згортки вхідного зображення для отримання карти ознак. Для виконання даної операції використовуються фільтри (матриця параметрів). Для даної нейронної мережі було обрано 4 рівня фільтрів по 32, 64, 128, 128 фільтрів на кожному рівні відповідно. Формування карти ознак можна здійснити за допомогою наступної формули:

$$M(i,j) = (K * X)(i,j) = \sum_m \sum_n K(m,n) X(i-m, j-n), \quad (1)$$

де M – елемент карти ознак з координатами i та j, X – вхідне зображення, K – детектор ознак, (m, n) – розмірності детектора ознак.

Етап 3 – Операція максимального об'єднання

Максимальне об'єднання є операцією, яка об'єднує елементи в межах фільтра на карті ознак і вибирає найбільший елемент. Тобто, після проходження через шар максимального об'єднання, отримується нова карта ознак, яка містить найбільш помітні ознаки з попередньої карти ознак. Виконати дану операцію можна за допомогою наступної формули:

$$p(i,j) = \max_{i,j}(x(i-m, j-n)), \quad (2)$$

де $p(i, j)$ – значення елементу поточного рівня з координатами i та j, x – вхідні дані з попередніх рівнів, (m, n) – розмірність рецептивного поля.

Етап 4 – Повнозв'язний рівень

Повнозв'язний рівень є моделлю багаторівневого персептрона, де всі нейрони з наступного шару з'єднані з нейронами попереднього шару. Цей рівень використовуються на передостанньому етапі роботи мережі для підготовки результатів на виході мережі. На даному рівні виконується обчислення скалярного добутку даних та параметрів з додаванням зсуву.

Етап 5 – Класифікація отриманих ознак за допомогою SVM

Метод опорних векторів (SVM) використовується для знаходження параметрів гіперплощини у багатомірному просторі, яка може служити для класифікації. Головна ідея полягає в тому, щоб знайти гіперплощину, яка має найбільшу відстань до найближчих навчальних точок будь-якого класу. Ця відстань називається «функціональним запасом». Чим більший функціональний запас, тим менша буде помилка узагальнення класифікатора. На виході класифікатор SVM видає оцінку, яка є ймовірністю того, що вхідні дані належать до певного класу (насильницького або не насильницького).

Алгоритмом, який оновлюватиме вагові коефіцієнти нейромережі під час процесу навчання обрано зворотне поширення похибки. Робота даного алгоритму полягає в наступному: відбувається процес знаходження градієнтів помилок – числових коефіцієнтів (відношення вхідних даних, зсуву до функції втрат), які використовуються для оновлення ваг кожного рівня мережі. Даний алгоритм здійснює оновлення ваг з кінця мережі до початку, у випадку архітектури, що розглядається, від повністю зв'язного рівня до рівня виконання операції згортки.

Для того, щоб запропонований метод міг працювати з відеоматеріалом в реальному часі під час процесу навчання нейронної мережі використовувався метод fine-tuning. Суть даного методу полягає в тому, що нейромережу спочатку навчають на наборі даних готових відео, а потім поступово додають до набору даних відеопотоки у реальному часі. Це допомагає нейромережі навчитися розпізнавати насильство в умовах шуму, швидкої зміни та різноманітності.

Аналіз ефективності запропонованого методу

Для того, щоб оцінити ефективність роботи запропонованого методу, проведено декілька експериментів щодо трьох наборів даних для виявлення насильства: «Hockey fights» [11], «Violent flows» [12], «Livestream» [13]. Оцінка ефективності методу базується на визначенні загальної точності для кожного набору даних.

Для визначення загальної точності роботи методу на відповідному набору даних застосовано наступний підхід:

Етап 1 – Визначення середньої точності

Необхідно зі всіх точностей (accuracy) знайти точність з найбільшим значенням, знаючи цю точність та її епоху потрібно сформувати новий масив даних точностей, які знаходяться в радіусі 10 епох від цієї максимальної точності. Отримавши масив точностей можна знайти середню точність за допомогою наступної формули:

$$A = \frac{\sum_{i=1}^{N} x_i}{N}, \tag{3}$$

де A – середня точність, N – загальна кількість влучень, x – значення відповідної точності, i – порядковий номер.

Етап 2 – Визначення стандартного відхилення

Отримавши середню точність можна знайти стандартне відхилення за наступною формулою:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N} (x_i - A)^2}{N}}, \tag{4}$$

де σ – стандартне відхилення, N – загальна кількість влучень, x – значення відповідної точності, i – порядковий номер, A – середня точність.

Таким чином, отримане значення середньої точності буде відповідати загальній точності, значення стандартного відхилення буде відповідати похибці середнього значення, яку можна виразити як ± значення, тобто загальна точність = середня точність ± стандартне відхилення.

Загальна точність для набору даних «Hockey fights» склала 98.5%, зображено на рисунку 2. Також у таблиці 1 наведено порівняння точності методів для набору даних «Hockey fights».

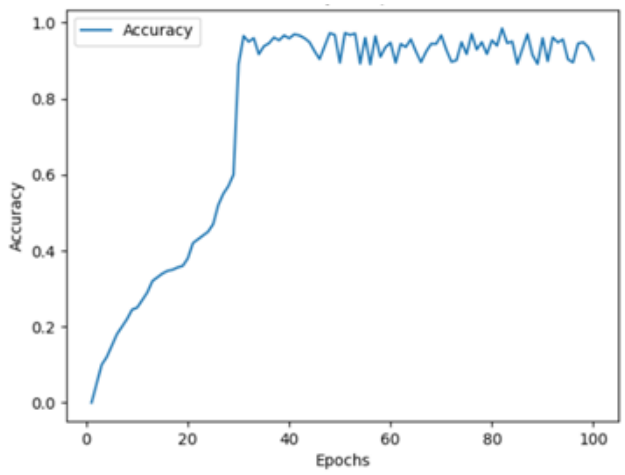

Рис. 2. Загальна точність набору даних «Hockey fights»

Загальна точність для набору даних «Violent flows» склала 99.45%, зображено на рисунку 3. Наведено у таблиці 2 порівняння точності методів для набору даних «Violent flows».

Таблиця 1

Порівняння точності методів на наборі даних «Hockey fights»

Метод	Hockey fights
Запропонований	98.5 ± 0.78%
Просторово-часовий кодер [7]	96.54 ± 1.01%
3D CNN [10]	98.3 ± 0.81%

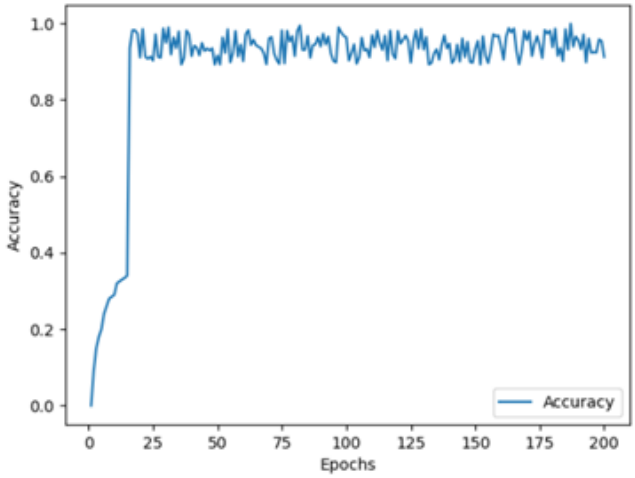

Рис. 3. Загальна точність набору даних «Violent flows»

Таблиця 2

Порівняння точності методів на наборі даних «Violent flows»

Метод	Violent flows
Запропонований	99.45 ± 0.37%
Просторово-часовий кодер [7]	92.18 ± 3.29%
3D CNN [10]	97.17 ± 0.95%

Загальна точність для набору даних «Livestream» склала 87.4%, зображено на рисунку 4. У таблиці 3 зображено порівняння точності методів для набора даних «Livestream».

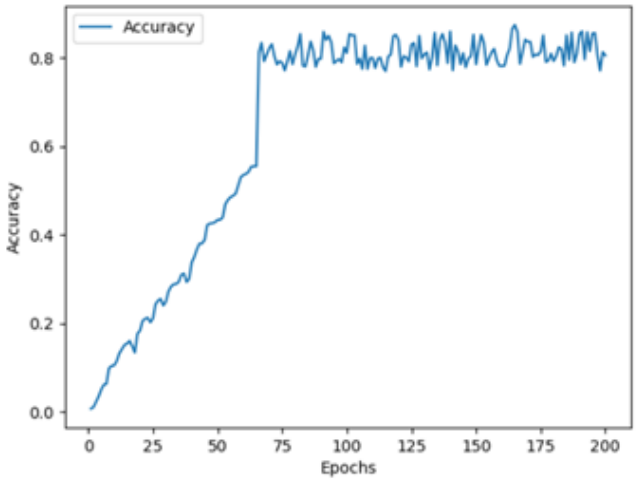

Рис. 4. Загальна точність набору даних «Livestream»

Таблиця 3

Порівняння точності методів на наборі даних «Livestream»

Метод	Livestream
Запропонований	87.4 ± 2.19%
Просторово-часовий кодер [7]	-
3D CNN [10]	-

Особливістю запропонованого методу є робота з відеоматеріалом (відеопотоком) у реальному часі. Дана можливість досягається за рахунок того, що згорткова нейронна мережа навчена на неперервному потоку даних з мультимедійних платформ для онлайн трансляцій використовуючи метод fine-tuning. Тобто, навчання відбувається в режимі реального часу і триватиме, доки примусово не зупиниться трансляція. Відповідно, тестування запропонованого методу відбувалося аналогічним чином, методу на вхід подавалася трансляція і він в реальному часі видавав оцінку сцени, яка відображалася на трансляції.

Висновки

Отже, запропонований метод для виявлення зовнішніх проявів насильства за допомогою нейромережевих засобів дозволяє визначити ступінь насильницького характеру у відсотковому відношенні, на статичних і динамічних відеоматеріалах. Метод на вхід приймає відеоматеріал з якого згорткова нейронна мережа вилучає набір ознак. Потім вилучений набір ознак передається класифікатору SVM, який визначає ймовірність належності вхідних даних до певного класу: насильницького або не насильницького. Головна особливість цього методу полягає в тому, що він може працювати з відеоматеріалом у реальному часі. Це досягається завдяки тому, що згорткова нейронна мережа навчалася на неперервному потоці даних із мультимедійних платформ для онлайн трансляцій за допомогою методу fine-tuning. Проведено експерименти з використанням різних наборів даних для оцінки ефективності запропонованого методу. Результати показали, що метод досягає високої точності (87,4%-99,45%) виявлення насильства та працює ефективно з відеопотоком даних у реальному часі.

Подальші дослідження спрямовані на пришвидшення роботи та покращення точності запропонованого методу для набору даних, які пов’язані з насильницькими діями у реальному часі.

References

1. Violence a global public health problem. https://www.scielo.br/j/csc/a/3hrn64cpBqBFb9mNfP4KGXr/?lang=en

2. Використання технологій штучного інтелекту протидії злочинності. https://ivpz.kh.ua/wp-content/uploads/2020/12/Матеріали-семінару_Використання-техн-штучного-інтел_5.11.2020.pdf

3. A CNN-RNN Combined Structure for Real-World Violence Detection in Surveillance Cameras. https://www.mdpi.com/2076-3417/12/3/1021

4. What Is a Convolutional Neural Network? https://www.ibm.com/topics/convolutional-neural-networks

5. Рекурентна нейронна мережа (RNN): види, навчання, приклади. https://neurohive.io/ru/osnovy-data-science/rekurrentnye-nejronnye-seti/

6. NABNet: A Nested Attention-guided BiConvLSTM network. https://www.sciencedirect.com/science/article/abs/pii/S1746809422007017

7. Convolutional LSTM for the Detection of Violence in Videos. https://openaccess.thecvf.com/content_ECCVW_2018/papers/11130/Hanson_Bidirectional_Convolutional_LSTM_for_the_Detection_of_Violence_in_Videos_ECCVW_2018_paper.pdf

8. Simonyan K., Zisserman A. Very deep convolutional networks for large-scale image recognition. In International Conference on Learning Representations (2015). http://arxiv.org/abs/1409.1556

9. 3D Convolutional Neural Network — A Guide for Engineers. https://www.neuralconcept.com/post/3d-convolutional-neural-network-a-guide-for-engineers

10. Jiang X., Xu K., Sun T., Li J. Efficient Violence Detection Using 3D Convolutional Neural Networks. https://www.researchgate.net/profile/Tanfeng-Sun/publication/337537845_Efficient_Violence_Detection_Using_3D_Convolutional_Neural_Networks/links/5f75e252a6fdcc00864ccb95/Efficient-Violence-Detection-Using-3D-Convolutional-Neural-Networks.pdf

11. Hockey Fight Detection Dataset. https://paperswithcode.com/dataset/hockey-fight-detection-dataset

12. Violent-Flows. https://paperswithcode.com/dataset/violent-flows

13. Livestream. https://www.twitch.tv/directory/category/just-chatting