



МЕТОДИ ВИЯВЛЕННЯ ОБЛИЧЧЯ ЛЮДИНИ НА ОСНОВІ КОМБІНОВАНОГО КАСКАДУ КЛАСИФІКАТОРІВ

Ігор Палій

Науково-дослідний інститут інтелектуальних комп'ютерних систем
 Тернопільський національний економічний університет
 пл. Перемоги 3, 46004, Тернопіль, Україна
 ipl@tneu.edu.ua, <http://www.tanet.edu.te.ua/ics/>

Анотація: у статті представлено покращений метод виявлення обличчя людини на основі комбінованого каскаду класифікаторів із вдосконаленим підходом до верифікації облич-кандидатів, а також методи та алгоритми генерації структури та навчання рівня верифікації каскаду (згорткової нейронної мережі). Комбінований каскад демонструє високий рівень виявлення при незначній кількості хибних позитивних виявлень, а запропонований підхід до верифікації кандидатів працює майже втричі швидше у порівнянні з класичною схемою верифікації. Метод генерації структури мережі дозволяє автоматично створити розріджену асиметричну структуру згорткової нейронної мережі. Покращений метод навчання використовує адаптивне співвідношення тренувальних прикладів для отримання навченої мережі з дуже низькою похибкою класифікації для позитивних прикладів.

Ключові слова: виявлення обличчя людини, комбінований каскад класифікаторів, каскад простих класифікаторів, згорткова нейронна мережа, верифікація кандидатів, алгоритм активного навчання.

1. ВСТУП

Першим кроком для будь-якої обробки обличчя людини є виявлення обличчя (ВО). Система ВО повинна визначити наявність або відсутність облич на вхідному зображенні, і, у випадку їхньої присутності, повернути координати та розмір кожного обличчя [1]. Багато зарубіжних (Сан, Поджіо, Роулі, Кенейд, Шнейдерман, Роз, Янг, Віола, Джонс, Лінхарт, Гарсія, Делакіс, Лі, Вежневец, Кудряшов, Фоменков, Самаль та ін.), а також вітчизняних вчених (Муригін, Ньюькін, Якименко, Дзюба, Антошук та ін.) внесли свій вклад у розвиток даного напрямку. Серед сфер застосування ВО можна виділити наступні: розпізнавання облич, пошук зображень за змістом, людино-комп'ютерний інтерфейс, аналіз емоцій, відеоконференція, відеоспостереження, підрахунок відвідувачів, контроль доступу та інші. Задача ВО є достатньо простою для людини, проте при створенні автоматичної системи ВО потрібно враховувати ряд факторів [1]: зовнішність, ракурс, нахил, вираз обличчя, наявність структурних елементів, перекриття іншими об'єктами, умови спостереження та ін.

Найкращі результати на сьогоднішній день

демонструють методи, що базуються на зовнішньому вигляді. Зокрема Сан і Поджіо [2] розробили перший точний метод виявлення фронтальних горизонтальних облич, що відноситься до даної групи. Авторами запропоновано ряд алгоритмів, які були використані у багатьох інших системах ВО: нормалізація вхідних зображень, генерація віртуальних позитивних прикладів (облич), самозагрузка негативних прикладів (необлич) у процесі активного навчання, стратегія пошуку облич по масштабу і розміщенню за допомогою фіксованого вікна класифікатора та ін.

Роулі та ін. [3] запропонували нейромережевий підхід, який продемонстрував хороші результати на великому і складному тестовому наборі MIT+CMU (130 зображень з 507 обличчями). Їхня система використовує нейронну мережу з групами локальних рецепторних полів. Кожне вхідне вікно зображення проходить попередню обробку, як і в [2]. Для зменшення кількості хибних позитивних виявлень (ХПВ) використано ансамбль нейронних мереж з різними стратегіями прийняття рішень. Швидка версія системи ВО має каскадну структуру і використовує додаткову нейронну мережу, яка сканує

зображення з більшим кроком. Знайдені кандидати верифікуються за допомогою базового класифікатора.

Дуже швидкий метод ВО представлений у [4], котрий використовує AdaBoost для вибору важливих Хаар-подібних рис і каскад простих класифікаторів (КПК).

Гарсія і Делакіс [5] запропонували новий підхід до ВО з використанням згорткової нейронної мережі (ЗНМ), яка розроблена Лекуном та ін. [6]. Їхня система автоматично в процесі навчання настроює оптимальні згорткові фільтри, що працюють в якості екстракторів рис. На відміну від попередніх методів, які сканують вхідне зображення за допомогою фіксованого вікна, ЗНМ обробляє зображення за раз на кожному рівні масштабування.

Сучасні методи ВО все ще мають ряд недоліків. Наприклад, системи на основі "сильних" монолітних класифікаторів [2, 3, 5] володіють низькою похибкою виявлення, дають мало ХПВ, але занадто повільні для роботи з відеопотоком. Навіть найшвидша система ВО даної групи з [5] потребує додатковий час для верифікації знайдених облич. Стадія верифікації у даному випадку пояснюється неточностями проєкції результатів виявлення з вихідного шару ЗНМ на вхідне зображення. Система ж на основі КПК [4] є дуже швидкою, але пропускає частково закриті або затінені обличчя, а також дає більше ХПВ.

У статті представлено покращений метод ВО на основі комбінованого каскаду класифікаторів [7], а також методи та алгоритми генерації структури та навчання ЗНМ. Крім того, подані висновки і майбутні напрямки досліджень.

2. МЕТОД ВИЯВЛЕННЯ ОБЛИЧЧЯ НА ОСНОВІ КОМБІНОВАНОГО КАСКАДУ КЛАСИФІКАТОРІВ

Каскадна структура класифікатора для ВО дозволяє досягти високої швидкості обробки вхідного зображення за рахунок швидкого відкидання більшості фону і зосередження уваги на областях, схожих на обличчя. Проте, у порівнянні з монолітними класифікаторами, збільшується похибка виявлення і рівень ХПВ [4]. Розширення набору Хаар-подібних рис [8, 9], вдосконалення алгоритму навчання [10] для КПК дозволили лише на 7-8% підвищити рівень виявлення при низьких рівнях ХПВ. Тому, замість того, щоб використовувати КПК як самостійний класифікатор для ВО, доцільно поєднати його з монолітним класифікатором в рамках дворівневого комбінованого каскаду класифікаторів (ККК) [7]. Перший рівень ККК,

представлений каскадом простих класифікаторів, відповідає за виявлення облич-кандидатів, а другий (ЗНМ) – за верифікацію знайдених кандидатів (рис. 1).

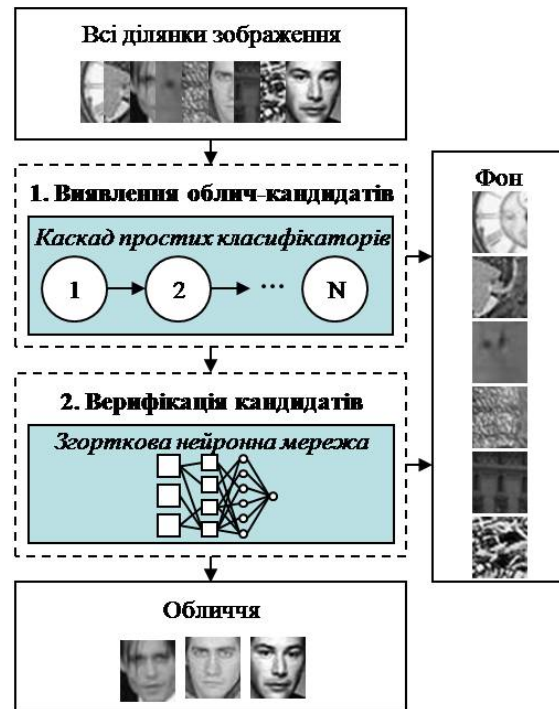


Рис. 1 – Процес виявлення обличчя на основі комбінованого каскаду класифікаторів.

Вибір КПК [4] в якості першого рівня комбінованого каскаду пояснюється тим, що він на сьогоднішній день дозволяє найшвидше виділяти обличчя-кандидати. КПК складається з рівнів, кожний з яких містить один і більше простих класифікаторів. Входом для кожного простого класифікатора є Хаар-подібна риса прямокутної форми, яка складається з "білих" і "чорних" прямокутників. Значення риси [4]

$$f(x) = w_w \times S_w + w_b \times S_b, \quad (1)$$

де x – область вхідного зображення, w_w та w_b – ваги для всього прямокутника риси і для його чорної частини, S_w S_b – суми пікселів всього прямокутника риси і його чорної частини (дані суми швидко обчислюються на основі інтегрального зображення [4]).

Вихідне значення простого класифікатора

$$h(x) = \begin{cases} 1, & \text{якщо } p \times f(x) < p \times \theta \\ -1, & \text{в іншому випадку} \end{cases}, \quad (2)$$

де p – полярність, яка показує напрям знаку нерівності, θ – порогове значення простого

класифікатора.

КПК являє собою лінійну комбінацію простих класифікаторів

$$H(x) = \text{sign}\left(\sum_{t=1}^T w_t \times h_t(x)\right), \quad (3)$$

де T – кількість простих класифікаторів, w_t – вага t -простого класифікатора. Вага простого класифікатора визначається протягом навчання КПК за допомогою алгоритму Gentle AdaBoost [11]

$$w_t = 0.5 \times \ln \frac{1 - \gamma_t}{\gamma_t}, \quad (4)$$

де γ_t – помилка t -простого класифікатора.

Вибір ЗНМ для верифікації знайдених обличчя-кандидатів пояснюється тим, що дана мережа в задачах класифікації, в силу особливостей своєї архітектури, володіє більшою стійкістю до деформацій (зсувів, зміни масштабу, поворотів, зашумленості) вхідних образів, ніж інші відомі класифікатори [6]. З метою пришвидшення роботи ЗНМ використано розріджену структуру мережі, а не повнозв'язну, а також зменшено кількість шарів шляхом здійснення площинами і згортки, і підвибірки одночасно [12] (рис. 2).

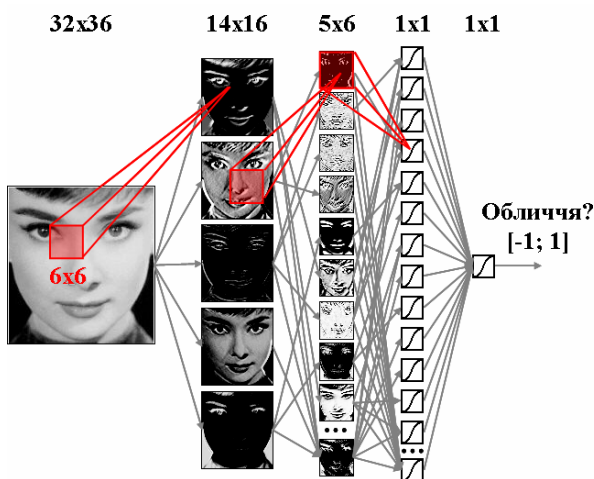


Рис. 2 – Структура згорткової нейронної мережі.

Вихідне значення нейрона з координатами (m,n) p -площини l -шару

$$y_{m,n}^{l,p}(x) = F(S_{m,n}^{l,p}(x)), \quad (5)$$

де x – вхідне зображення обличчя-кандидата, F – функція активації, $S_{m,n}^{l,p}(x)$ – зважена сума нейрона.

В якості функції активації в ЗНМ використано біполярну сигмоїдну функцію з вихідним діапазоном $[-1; 1]$, що рекомендовано в [13]. Тому вираз (5) набуває наступного вигляду:

$$y_{m,n}^{l,p}(x) = \frac{2}{1 + \exp(-S_{m,n}^{l,p}(x))} - 1, \quad (6)$$

в якому зважена сума

$$S_{m,n}^{l,p}(x) = \left(\sum_{k=0}^{K-1} \sum_{r=0}^{R-1} \sum_{c=0}^{C-1} y_{2m+r, 2n+c}^{l-1,k}(x) \times w_{r,c}^{l,p,k}\right) - b^{l,p}, \quad (7)$$

де K – кількість вхідних площин (а також і ядер згортки), R та C – висота і ширина ядра згортки, $w_{r,c}^{l,p,k}$ – синаптична вага з координатами (r, c) у ядрі згортки між k -площиною $(l-1)$ -шару і p -площиною l -шару, $b^{l,p}$ – порогове значення нейронів p -площини l -шару, $y_{2m+r, 2n+c}^{l-1,k}(x)$ – вихідне значення $(l-1)$ -шару k -площини з координатами $(2m+r, 2n+c)$.

Так як КПК використовується не самостійно, а в рамках ККК, то для збільшення потенційного рівня виявлення порогове значення мінімальної кількості багатократних виявлень, необхідних для прийняття області зображення як обличчя, зменшено до мінімуму (до 1). Логічно, що разом з цим серед кандидатів зростає і кількість ХПВ (рівень виявлення 89.9% при 288 ХПВ на тестовому наборі МІТ+СМУ). Проте застосування ЗНМ дозволяє відсіяти складні ХПВ і прийняти майже всі обличчя (88.4% і 8).

Верифікацію кожного знайденого обличчя-кандидата можна здійснити за двома підходами: 1) шляхом прямої його подачі на вхід ЗНМ (рис. 3а), 2) шляхом сканування фіксованим вікном розширеної області зображення кандидата в сусідніх позиціях при декількох масштабах і подачею кожного вікна на вхід ЗНМ [3, 5] (рис. 3б). Проведені автором дослідження першого підходу до верифікації [7] показали, що він є швидким і простим у реалізації (кількість симуляцій ЗНМ рівна кількості кандидатів), але дозволяє отримати дещо гірші результати, ніж у відомих методів ВО: на тестовому наборі МІТ+СМУ досягнуто рівня виявлення 75.4% при 8 ХПВ, – тому надалі даний підхід не розглядається.

Другий підхід дозволяє приймати обличчя-кандидати, навіть при неточному виявленні їхніх координат по масштабу чи розміщенню, але характеризується вищою обчислювальною складністю, так як збільшується кількість симуляцій ЗНМ для одного обличчя-кандидата.

Сповільнення роботи ККК більш ніж у 4 рази у порівнянні з КПК є критичним при роботі з відеопотоком, тому даний підхід теж є неприйнятним.

Автором запропоновано застосувати підхід до верифікації (рис. 3в), який використовує можливість ЗНМ обробляти вхідне зображення будь-якого розміру за раз, що зберігає переваги другого підходу, значно зменшуючи при цьому обчислювальну складність (табл. 1).

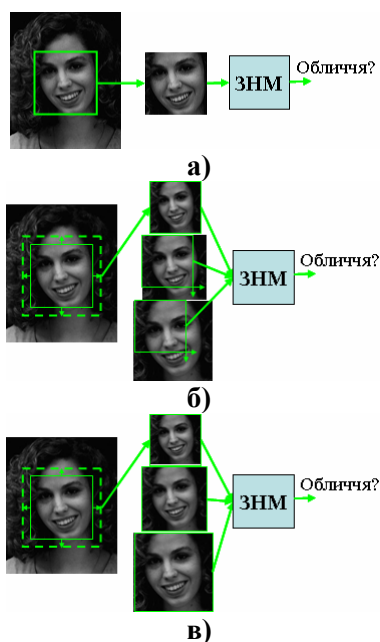


Рис. 3 – Верифікація облич-кандидатів.

Таблиця 1. Порівняння обчислювальної складності підходів до верифікації на тестовому наборі MIT+CMU

Підхід до верифікації	Кількість симуляцій ЗНМ, шт.	Загальний час ВО, сек.
Другий	92401	121
Запропонований	3737	37

При порівнянні рівнів ВО другого та запропонованого підходів до верифікації (рис. 4) виявилось, що при використанні другого підходу комбінований каскад демонструє дещо вищі показники рівня ВО (максимум на 3%) при однакових кількостях ХПВ, проте запропонований підхід демонструє більш ніж трьохкратну перевагу у швидкодії (див. табл. 1).

Для подальшого пришвидшення верифікації запропоновано використати механізм прийняття облич, який полягає у тому, що кандидат приймається як обличчя, при досягненні порогового значення кількості багатократних виявлень. При цьому подальша верифікація кандидата зупиняється, а отже і зменшується кількість симуляцій ЗНМ.

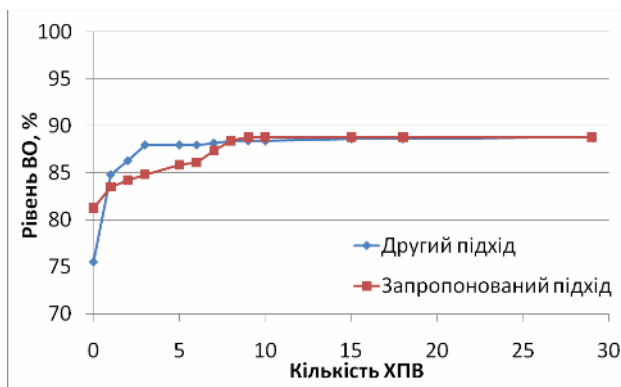


Рис. 4 – ROC-криві комбінованого каскаду класифікаторів на тестовому наборі MIT+CMU.

На рис. 5 зображено алгоритм роботи запропонованого підходу до верифікації з механізмом прийняття облич, де FC – кількість облич-кандидатів, $SCALES$ – кількість зображень піраміди по масштабу для f -кандидата, MD – кількість багатократних виявлень, y – вихідне значення ЗНМ, Min_MD – порогове значення багатократних виявлень.

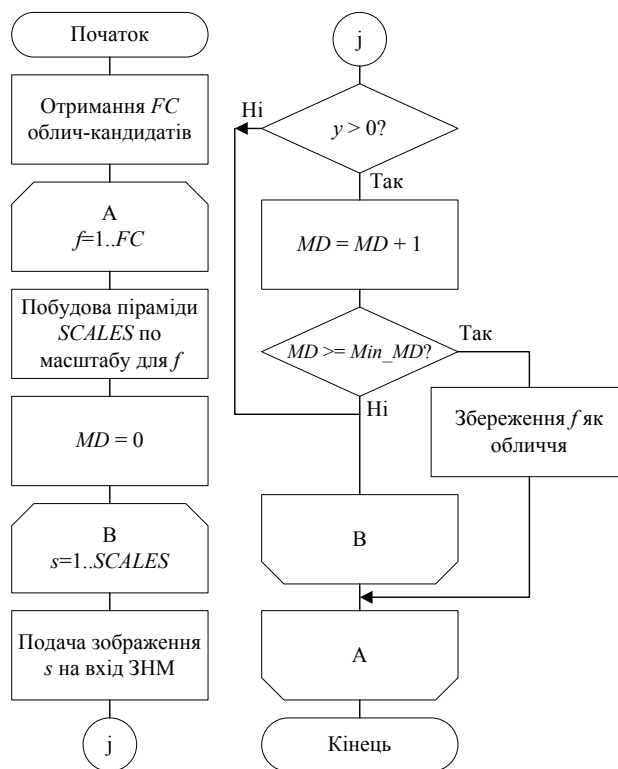


Рис. 5 – Алгоритм роботи запропонованого підходу до верифікації кандидатів.

Метод на основі ККК із запропонованим підходом до верифікації кандидатів демонструє одні з найкращих результатів у порівнянні з існуючими системами ВО (табл. 2.).

Таблиця 2. Порівняння результатів роботи методів ВО на тестовому наборі MIT+CMU

Метод ВО	Рівень ВО, %	Кількість ХПВ, шт.
Роулі та ін. [3]	83.2	10
Роулі та ін. [3] (швидка версія)	76.9	8
Віола і Джонс [4]	76.1	10
Лі та ін. [10]	83.6	10
Гарсія та Делакіс [5]	90.5	10
Ферод [14]	86.0	8
Кудряшов [9]	83.0	7
ККК (КПК+ЗНМ)	88.4	8

Як видно з таблиці 2, за рівнем виявлення при однаковій кількості ХПВ комбінований каскад поступається приблизно на 2% тільки методу Гарсія та Делакіс [5], але при цьому ККК обробляє зображення більш ніж у 8 разів швидше за систему з [5].

Програмна реалізація системи виявлення обличчя на основі розробленого методу здійснена у середовищі Visual C++ 2007 з використанням бібліотеки Intel OpenCV (Open Computer Vision Library) [15] та Intel IPP (Integrated Performance Primitives) [16], що дозволяє підвищити швидкодію системи на робочих станціях з Intel-процесорами. Розроблену систему протестовано в НДІ інтелектуальних комп'ютерних систем з використанням камери Logitech QuickCam Messenger на робочій станції Intel Celeron 2.3GHz. При використанні параметрів по замовчуванню система обробляє вхідне зображення розміром 352x288 пікселів зі швидкістю 8-10 к./сек.

3. МЕТОД АВТОМАТИЧНОЇ ГЕНЕРАЦІЇ СТРУКТУРИ ЗГОРТКОВОЇ НЕЙРОННОЇ МЕРЕЖІ

Існує декілька стратегій зв'язування площин між шарами ЗНМ: повна, Тепліца, бінарна і задана вручну [17]. Легко програмуються всі стратегії окрім останньої, що задається вручну відповідною таблицею зв'язків між площинами попереднього і поточного шарів. Проте тільки остання дозволяє створити несиметричну зв'язність структури ЗНМ, як це рекомендується у [6]. Зв'язність, задана вручну, не є універсальною стратегією, так як при зміні кількості шарів, площин потрібно вручну оновлювати таблицю зв'язків. Тому автором запропоновано метод для автоматизації створення структури ЗНМ з використанням генератора випадкових чисел.

Шари та площини ЗНМ створюються

автоматично, на основі наступних вхідних параметрів: кількість площин вхідного шару, розмір тренувального зображення, розмір та крок ядра згортки, коефіцієнт площин (визначає кількість площин кожного наступного шару). Розмір зображення для навчання, розмір та крок ядра згортки підбираються таким чином, щоб згідно (8) і (9) отримати цілі значення розмірів площини для кожного наступного шару, а також отримати площину розміром 1x1 нейрон для вихідного шару.

$$h^{l+1} = (h^l - h_k^{l+1}) / s_{k,r}^{l+1} + 1, \quad (8)$$

$$w^{l+1} = (w^l - w_k^{l+1}) / s_{k,c}^{l+1} + 1, \quad (9)$$

де h^l та w^l – висота та ширина площини l -шару, h_k^{l+1} та w_k^{l+1} – висота та ширина ядра згортки $(l+1)$ -шару, $s_{k,r}^{l+1}$ та $s_{k,c}^{l+1}$ – крок ядра згортки $(l+1)$ -шару по вертикалі та горизонталі.

У даній роботі розмір тренувального зображення для ЗНМ рівний 32x36 пікселів, ядра згортки – 6x6 пікселів, крок ядра по рядках та стовпчиках рівний 2. Кількість площин вхідного шару та коефіцієнт площин вибрано експериментально (5 і 2 відповідно).

Кількість площин для кожного наступного шару (крім останнього)

$$p^{l+1} = p^l \times PK, \quad (10)$$

де PK – коефіцієнт площин.

Зв'язування площин здійснюється після створення всіх шарів та площин і передбачає обчислення для площини будь-якого шару (крім останньої) значення кількості зв'язків

$$c^{l,p} = Rand(1..p^{l-1}) \times CK, \quad (11)$$

де $Rand$ – функція генерації випадкових цілих чисел за рівномірним законом розподілу, CK – коефіцієнт зв'язаності.

При зв'язуванні площин враховуються наступні обмеження:

- кількість зв'язків площини не може перевищувати кількість площин попереднього шару;
- кількість зв'язків площини повинна бути більшою за 0.

Для кожного зв'язку випадково вибирається номер площини попереднього шару. Остання площина будь-якого шару, а також єдина площина вихідного шару, зв'язана зі всіма площинами попереднього шару. Це дозволяє

уникнути появи незв'язаних площин, коли вихідні значення площини не подаються на площини наступного шару. Коефіцієнт зв'язаності дозволяє змінювати ступінь розрідженості мережі (у роботі використано $CK=0.25$, який вибрано експериментально).

Використання більших значень розміру ядра згортки, кроку ядра у порівнянні з мережею у [5, 6], а також малого коефіцієнту зв'язаності зумовлено необхідністю зменшення кількості шарів і зв'язків для генерації швидкої ЗНМ.

4. МЕТОД НАВЧАННЯ ЗГОРТКОВОЇ НЕЙРОННОЇ МЕРЕЖІ

Існують два підходи до організації навчання нейронних мереж: всеохоплююче і активне навчання [3]. Всеохоплююче навчання можна застосувати, коли в наявності є тренувальні приклади всіх класів, і розмір тренувального набору не є надто великим. Для ефективного навчання класифікатора облич/необлич потрібні десятки тисяч позитивних тренувальних прикладів, а кількість негативних тренувальних прикладів практично необмежена. Тому доцільно використати активне навчання, яке дозволяє тренувати класифікатор на відносно невеликому навчальному наборі і ефективно відбирати негативні приклади.

Для тренування використано 20113 позитивних прикладів та близько 350 зображень, які не містять облич, для набору негативних прикладів у процесі виконання процедури самозагрузки (bootstrapping) [2, 3, 5]. Крім того, сформовано оціночний набір з 1000 позитивних прикладів, що не входять в тренувальний набір, та 1000 негативних.

Так як ЗНМ використовується не самостійно, а в рамках ККК, який на першому рівні відкидає більшість фону, то більша увага повинна приділятися досягненню низької похибки класифікації на позитивних прикладах, а не негативних. При використанні активного навчання невідомо наперед, на яких саме тестових прикладах (позитивних чи негативних)

навчена ЗНМ буде давати більшу похибку. Тому автором покращено метод активного навчання шляхом застосування адаптивного співвідношення між позитивними та негативними прикладами, які на кожній ітерації навчання додаються до активного тренувального набору. Співвідношення для позитивних/негативних тренувальних прикладів

$$PER = \begin{cases} PER - 0.1 \text{ якщо } PEE \leq 1\% \\ PER + 0.1 \text{ якщо } PEE > 1\% \end{cases}, \quad (12)$$

де PEE – відносна похибка класифікації на оціночному наборі позитивних прикладів. Порогове значення для PEE , як і крок зміни PER вибрано експериментально.

Безпосереднє навчання ЗНМ здійснюється за допомогою послідовного алгоритму зворотнього розповсюдження помилки з адаптивним кроком навчання [18], так як даний алгоритм забезпечує більшу швидкість і стабільність навчання у порівнянні з пакетними алгоритмами [19]. Згідно алгоритму, модифікація значень синаптичних ваг та порогів на кожній ітерації (після подачі кожного тренувального образу) відбувається на основі наступних формул:

$$\Delta w_{r,c}^{l,p,k} = -\alpha^{l,p} \times \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \gamma_{m,n}^{l,p} \times F'(S_{m,n}^{l,p}) \times y_{2m+r,2n+c}^{l-1,k}, \quad (13)$$

$$\Delta b^{l,p} = \alpha^{l,p} \times \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \gamma_{m,n}^{l,p} \times F'(S_{m,n}^{l,p}), \quad (14)$$

де $\Delta w_{r,c}^{l,p,k}$ та $\Delta b^{l,p}$ – зміни синаптичних ваг та порогів, $\alpha^{l,p}$ – адаптивний крок навчання для p -площини l -шару, який розраховується за допомогою (15), M та N – висота і ширина p -площини l -шару, $\gamma_{m,n}^{l,p}$ – помилка нейрону з координатами (m, n) на p -площині l -шару, яка для прихованого і вихідного шарів обчислюється за допомогою (16) та (17) відповідно.

$$\alpha^{l,p} = \frac{\sum_{m=0}^{M-1} \sum_{n=0}^{N-1} (\gamma_{m,n}^{l,p})^2 \times F'(S_{m,n}^{l,p})}{F'(0) \times (1 + \sum_{k=0}^{K-1} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \sum_{r=0}^{R-1} \sum_{c=0}^{C-1} (y_{2m+r,2n+c}^{l-1,k})^2) \times \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} (\gamma_{m,n}^{l,p})^2 \times (F'(S_{m,n}^{l,p}))^2}, \quad (15)$$

$$\gamma_{m,n}^{l,p} = \sum_{q=0}^{Q-1} \sum_{r=0}^{R-1} \sum_{c=0}^{C-1} \gamma_{2r+shR,2c+shC}^{l+1,q} \times F'(S_{2r+shR,2c+shC}^{l+1,q}) \times w_{2r,2c}^{l+1,q,p}, \quad (16)$$

$$\gamma = y - d, \quad (17)$$

де Q – кількість площин $(l+1)$ -шару, shR та shC – зсув по рядкам і стовпчикам для визначення позиції першого нейрона $(l+1)$ -шару, помилка якого впливає на помилку нейрона з координатами (m, n) l -шару, d – еталонне вихідне значення.

Похідна вихідного значення нейрона

$$F'(S_{m,n}^{l,p}) = 0.5 \times (1 - (y_{m,n}^{l,p})^2). \quad (18)$$

Згідно [16] абсолютне значення адаптивного кроку навчання обмежується значенням 10. Крім того, при його обчисленні доцільно проводити перевірку, щоб знаменник не дорівнював нулю.

Модифікація синаптичних ваг та порогів направлена на мінімізацію на кожній епосі навчання (після подачі всіх образів тренувального набору) сумарної середньоквадратичної помилки ЗНМ

$$E = 0.5 \times \sum_{x=0}^{X-1} \gamma_x^2, \quad (19)$$

де γ_x – помилка нейрона вихідного шару ЗНМ для вхідного образу x .

Алгоритм реалізації покращеного методу навчання ЗНМ показано на рис. 6, де W та B – синаптичні ваги та порогові значення ЗНМ, $MinY$ – мінімальне вихідне значення ЗНМ, $ActTSet$ – активний тренувальний набір, PE та NE – позитивний та негативний приклад, $PETSet$ – тренувальний набір позитивних прикладів, $BImgs$ – набір зображень, які не містять обличчя, $ActEp$ та $MaxActEp$ – поточна та задана максимальна епоха активного навчання, $MinE$ – задана мінімальна сумарна середньоквадратична помилка мережі, $MaxEp$ – задана максимальна епоха навчання, $VSet$ – оціночний набір. Навчання ЗНМ триває 12-15 епох активного навчання.

Експериментальні дослідження показали, що при навчанні згорткової нейронної мережі з фіксованим співвідношенням тренувальних прикладів ($PER=0.5$) відносна похибка класифікації для позитивних прикладів оціночного набору на останніх епохах починає зростати (рис. 7а). При використанні покращеного методу відносна похибка для позитивних прикладів при навчанні зберігається приблизно на рівні 1% (рис. 7б).

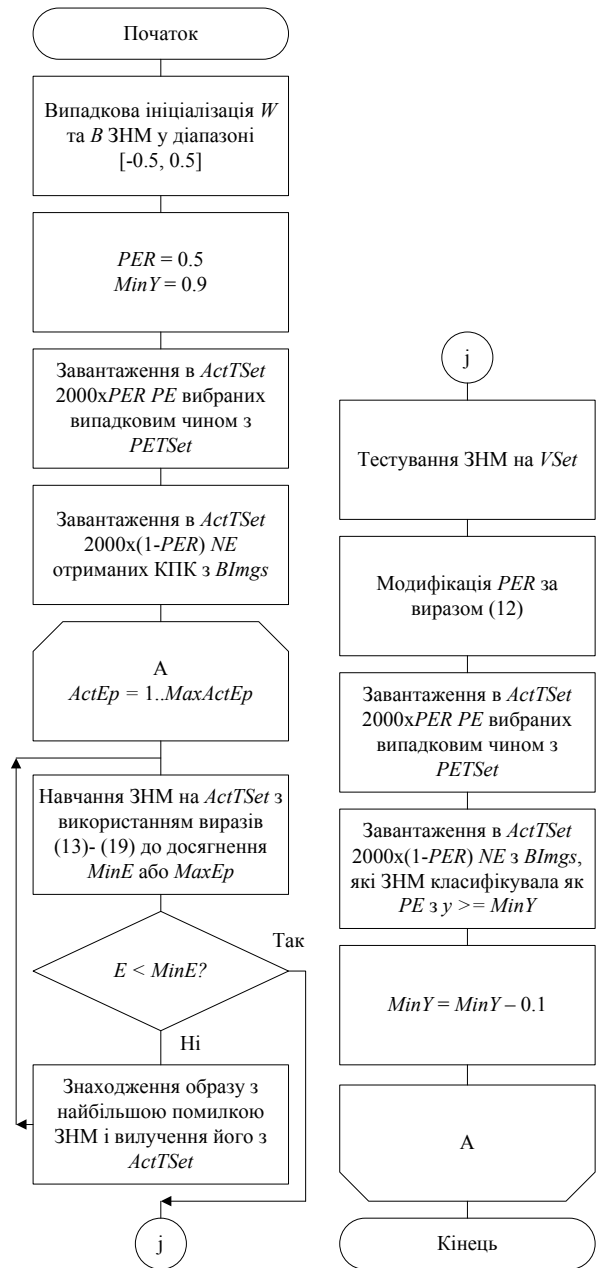
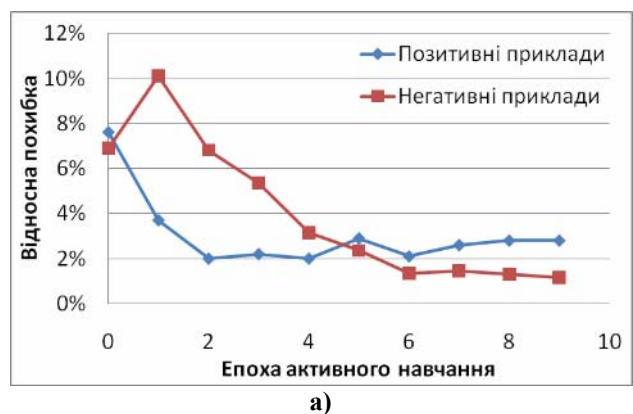


Рис. 6 – Алгоритм реалізації покращеного методу навчання згорткової нейронної мережі.



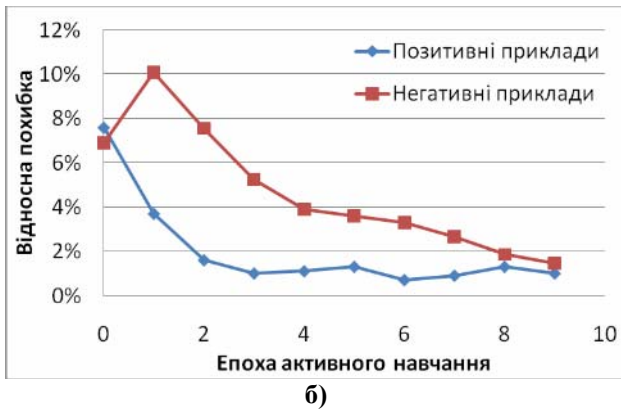


Рис. 7 – Порівняння результатів роботи відомого (а) та покращеного (б) методів активного навчання нейронної мережі на оціночному наборі.

5. ВИСНОВКИ І МАЙБУТНІ НАПРЯМКИ ДОСЛІДЖЕНЬ

Вдосконалено метод виявлення обличчя людини шляхом використання комбінованого каскаду класифікаторів та запропонованого підходу до верифікації обличч-кандидатів. Вдосконалений метод дозволяє підвищити рівень виявлення обличчя на 12% у порівнянні з відомим методом, що базується на каскаді простих класифікаторів, а також він є у 8 разів швидший за відомий метод на основі згорткової нейронної мережі.

Розроблено покращений метод навчання згорткової нейронної мережі, що дозволяє отримати навчену мережу з відносною похибкою класифікації на рівні 1% для позитивних прикладів оціночного набору.

Перспективними є наступні напрямки майбутніх досліджень:

- розпаралелення навчання комбінованого каскаду класифікаторів на обчислювальному кластері;
- пришвидшення роботи системи виявлення обличчя за допомогою алгоритмів сегментації за рухом та кольором шкіри [20].

6. СПИСОК ЛІТЕРАТУРИ

- [1] M. H. Yang. Recent Advances in Face Detection. *IEEE ICPR 2004 Tutorial*, Cambridge, United Kingdom, 2004, 93 p.
- [2] K. K. Sung and T. Poggio. Example-based learning for view-based human face detection // *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no.1, 1998, pp.39-51.
- [3] H. Rowley, S. Baluja, and T. Kanade. Neural network-based face detection // *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.20, 1998, pp.22–38.
- [4] P. Viola, M. Jones. Robust Real-Time Face Detection // *International Journal of Computer Vision*, 57 (2), 2004, pp. 137–154.
- [5] C. Garcia, M. Delakis. A Neural Architecture for Fast and Robust Face Detection // *IEEE-IAPR International Conference on Pattern Recognition (ICPR2002)*, Québec City, 2002, pp. 40-43.
- [6] Y. LeCun, L. Bottou, Y. Bengio and etc. Gradient-Based Learning Applied to Document Recognition // *Intelligent Signal Processing, IEEE Press*, 2001, pp. 306-351.
- [7] I. Paliy, A. Sachenko, Y. Kurylyak, V. Turchenko, V. Kapura. Face Detection Using Combined Cascade of Classifiers for Video Surveillance // *Scientific journal of the Brest State Technical University: physics, mathematics and informatics series*, 2007, pp. 5-8 (in Russian).
- [8] R. Lienhart, J. Maydt. An Extended Set of Haar-like Features for Rapid Object Detection // *IEEE ICIP 2002*, Vol. 1, Sep. 2002, pp. 900-903.
- [9] P. Kudryashov. Human Face Detection Algorithms for Image Analysis and Processing Applications. *Abstract of PhD. Thesis, Volgograd*, 2007, 23 p. (in Russian).
- [10] S. Z. Li, Z. Q. Zhang. FloatBoost Learning and Statistical Face Detection // *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, No. 9, September 2004, pp. 1112-1123.
- [11] Y. Freund, R. Schapire. *A decision-theoretic generalization of on-line learning and an application to boosting*. In *Computational Learning Theory: Eurocolt 95*, Springer-Verlag, 1995, pp. 23–37.
- [12] P. Simard, D. Steinkraus, J. Platt. Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis // *Seventh International Conference on Document Analysis and Recognition (ICDAR'03)*, Vol. 2, 2003, p. 958.
- [13] A.I. Wasserman. *Neural Computing: Theory and Practice*. – New York: Van Nostrand Reinhold, 1989, 230 p.
- [14] R. Feraud, O. Bernier, J. Viallet, M. Collobert. A Fast and Accurate Face Detector Based on Neural Networks // *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no.1, 2001, pp. 42-53
- [15] OpenCV library: <http://sourceforge.net/projects/opencv/>.
- [16] <http://www.intel.com/cd/software/products/asmo-na/eng/302910.htm>.
- [17] F. Tivive, A. Bouzerdoum. Efficient training algorithms for a class of shunting inhibitory

convolutional neural networks // *IEEE Transactions on Neural Networks*, May 2005, 16(3), pp. 541-556.

- [18] V. Golovko. *Neurointelligence: theory and application*. Brest, vol.1, 1999, 264 p.
- [19] V. Tsaregorodtsev. Total Inefficiency of Neural Network's Batch Training // *Proceedings of the XII Seminar on Neuroinformatics Applications*, Krasnoyarsk, 2004, pp. 145-151 (in Russian).
- [20] I. Paliy, Y. Kurylyak, A. Sachenko. Improved Neural Network-Based Face Detection Method // *Scientific journal of the Khmelnytskyi National University*, 2007, No 2, vol. 1, pp. 145-148 (in Ukrainian).
-



Ігор Палій. Науковий співробітник НДІ інтелектуальних комп'ютерних систем (НДІ ІКС), Тернопільського національного економічного університету (ТНЕУ), м. Тернопіль, Україна.

Отримав диплом спеціаліста з інформаційних систем в менеджменті в ТНЕУ (2002), диплом магістра з економічної кібернетики в ТНЕУ (2003). В даний час є аспірантом кафедри інформаційно-обчислювальних систем та управління факультету комп'ютерних інформаційних технологій, ТНЕУ.

Наукові інтереси: виявлення і розпізнавання облич, обробка зображень, штучні нейронні мережі, паралельні обчислення.



HUMAN FACE DETECTION METHODS USING COMBINED CASCADE OF CLASSIFIERS

Ihor Paliy

Research Institute of Intelligent Computer Systems
 Ternopil National Economic University
 3 Peremoga Square, 46004, Ternopil, Ukraine
 ipl@tneu.edu.ua, <http://www.tanet.edu.te.ua/ics/>

Abstract: *The paper presents the improved human face detection method using the combined cascade of classifiers with the improved face candidates' verification approach, as well as methods and algorithms for the verification level (convolutional neural network) structure generation and training. The combined cascade shows a high detection rate with a very small number of false positives and the proposed candidates' verification approach is in almost 3 times faster in comparison with the classic verification scheme. The network's structure generation method allows creating the sparse asymmetric structure of the convolutional neural network automatically. The improved training method uses the adaptive training examples ratio to obtain a trained network with a very low classification error for the positive examples.*

Keywords: *human face detection, combined cascade of classifiers, cascade of weak classifiers, convolutional neural network, candidates' verification, active training algorithm.*

1. INTRODUCTION

Human face detection (FD) is always the first step to any face processing, and it is also a challenging issue because of the great facial variability [1]. Recent FD methods from appearance-based group [2, 3, 4] show excellent results on MIT+CMU benchmark test set (130 images with 507 faces) [2], but they still have some disadvantages. For example, the FD systems with "strong" monolithic classifiers [2, 4] have a low detection error and false positives (FP) rate but they are too slow for the real-time video flow processing. The FD system based on the cascade of "weak" classifiers [3] is very fast but gives more FP.

The improved FD method based on the combined cascade of classifiers (CCC) [5], which overcome the mentioned issues, is presented in the paper as well as the methods for the automatic verification level's structure generation and active training.

2. FACE DETECTION METHOD USING COMBINED CASCADE OF CLASSIFIERS

The cascade structure of the FD classifier allows achieving higher image processing speed in comparison with the monolithic classifier architecture because of the fast background rejection [2, 3]. At the same time the cascade structure

increases the detection error and FP rate [3]. Therefore, it is proposed to use CCC which consists of two stages: face candidates' detection and candidates' verification. The first stage uses the cascade of weak classifiers [3], which is a linear combination of the weak classifiers

$$H(x) = \text{sign}\left(\sum_{t=1}^T w_t \times h_t(x)\right), \quad (1)$$

where x – input image area, T – the number of weak classifiers, w_t – t -weak classifier's weight value, $h_t(x)$ – t -weak classifier's output value.

The verification stage uses the convolutional neural network (CNN) [4] which due to the architecture features is the most robust to the input images' deformations (shifts, scales, rotations, occlusions) [6]. The network uses bipolar sigmoid activation function and the CNN's output value

$$y_{m,n}^{l,p}(x) = \frac{2}{1 + \exp(-S_{m,n}^{l,p}(x))} - 1, \quad (2)$$

where x – input face candidate's image, $S_{m,n}^{l,p}(x)$ – weighted sum of the neuron with coordinates (m,n)

on the p -plane of the l -layer.

The verification of each face candidate may be performed using two approaches: 1) straight verification of candidate's area by CNN [3], 2) extended candidate's area scanning by CNN using fixed window at several scale levels [4]. Experimental researches showed that the first approach provides a low detection rate [5]. The second one is too slow due to the large number on CNN simulations for each candidate. Therefore, the author proposed the verification approach which uses the CNN property to process the extended candidate's area at once on every scale level. The second and proposed verification approaches have almost the same detection and FP rates (Fig. 1), but the proposed one is in more than 3 times faster than the previous.

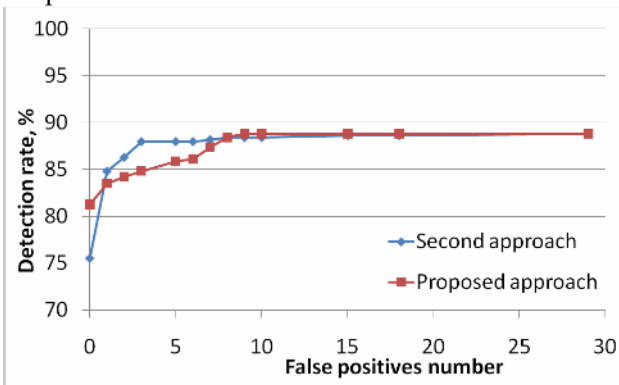


Fig. 1 – ROC-curves of combined cascade of classifiers on MIT+CMU test set.

The proposed verification approach also includes the face candidates' acceptance mechanism which allows decreasing the number of CNN simulations.

The improved FD method yields only to the system of Garcia et al. (Table 1), but the combined cascade of classifiers is approximately in 8 times faster.

Table 1. Face detection methods' results comparison using MIT+CMU test set

FD method	FD rate, %	FP number
Rowley et al. [2]	83.2	10
Rowley et al. [2] (fast version)	76.9	8
Viola and Jones [3]	76.1	10
Li et al. [7]	83.6	10
Garcia and Delakis [4]	90.5	10
Feraud [8]	86.0	8
Kudryashov [9]	83.0	7
Improved FD method	88.4	8

3. CONVOLUTIONAL NEURAL NETWORK'S STRUCTURE GENERATION METHOD

It is proposed to automatically generate sparse

asymmetric CNN structure which is more flexible than other strategies: binary, Toeplitz, manual. The structure creation process depends on the following input parameters: training image size, convolutional kernel size and steps, input layer planes' number and planes' coefficient. The training image size, kernel size and steps are chosen in order to provide the integer values of planes' sizes calculated using (3) and (4), and also to achieve the output layer plane's size equal to 1x1 neurons.

$$h^{l+1} = (h^l - h_k^{l+1}) / s_{k,r}^{l+1} + 1, \quad (3)$$

$$w^{l+1} = (w^l - w_k^{l+1}) / s_{k,c}^{l+1} + 1, \quad (4)$$

where h^l and w^l – l -layer plane's height and width, h_k^{l+1} and w_k^{l+1} – $(l+1)$ -layer convolutional kernel's height and width, $s_{k,r}^{l+1}$ and $s_{k,c}^{l+1}$ – $(l+1)$ -layer convolutional kernel's steps across rows and columns.

The planes' number for the next layer

$$p^{l+1} = p^l \times PK, \quad (5)$$

where PK – planes' coefficient.

The random-number generator is used to calculate the planes' number for each layer (except the last one). The input planes for the connections are chosen randomly too. The last plane of each layer and the output layer's plane are connected with all planes of the previous layer.

The sparse asymmetric structure is generated according to the recommendations in [6] and in order to speed up the CNN simulation time.

4. CONVOLUTIONAL NEURAL NETWORK TRAINING METHOD

Taking into account that CNN is used in the frame of CCC it's necessary to pay more attention to the classification error on the positive examples (faces) than on the negative ones (nonfaces). The active training method used in [2, 4] makes it impossible to predict the positive examples error trend. Therefore, the author improved the active training method by using the adaptive positive/negative training examples ratio which depends on the CNN validation set error on each active training epoch

$$PER = \begin{cases} PER - 0.1 \text{ якщо } PEE \leq 1\% \\ PER + 0.1 \text{ якщо } PEE > 1\% \end{cases}, \quad (6)$$

where PEE – relative positive examples' error on the validation set.

On-line gradient backpropagation with an adaptive training step [10] is used to calculate the modifications of CNN synapses' weights and biases on each active training epoch

$$\Delta w_{r,c}^{l,p,k} = -\alpha^{l,p} \times \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \gamma_{m,n}^{l,p} \times F'(S_{m,n}^{l,p}) \times y_{2m+r,2n+c}^{l-1,k}, \quad (7)$$

$$\Delta b^{l,p} = \alpha^{l,p} \times \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \gamma_{m,n}^{l,p} \times F'(S_{m,n}^{l,p}), \quad (8)$$

where $\alpha^{l,p}$ – adaptive training step for p -plane of l -layer, M and N – height and width of p -plane of l -layer, $\gamma_{m,n}^{l,p}$ – error of the neuron with coordinates (m, n) on p -plane of l -layer, $F'(S_{m,n}^{l,p})$ – derivative of the neuron's output value.

Experimental researches showed that the positive examples error on the validation set gradually grows on the last epochs when using the known training method [2, 4] with the fixed $PER=0.5$ (Fig. 2a). The improved method allows preserving low relative positive examples error which is equal to approximately 1% (Fig. 2b).

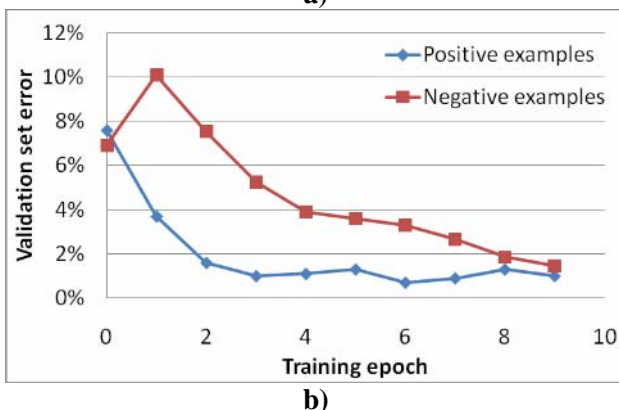
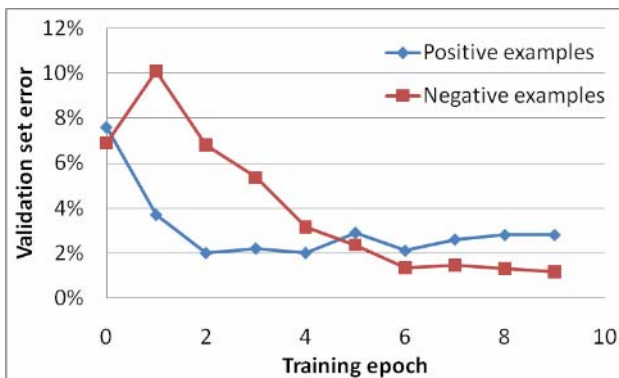


Fig. 2 – Comparison of known (a) and improved (b) active training methods results.

5. CONCLUSIONS AND FUTURE RESEARCHES

Human face detection method is improved using

the combined cascade of classifiers and proposed candidates' verification approach. The improved method allows increasing the detection rate on 12% in comparison with method from [3] and it is also in 8 times faster than the method in [4].

The improved active training method for the convolutional neural network is developed. It allows preserving a very low positive examples' error of approximately 1% on the validation set during the network training.

Promising future research directions include the parallelization of the combined cascade of classifiers training using the computing cluster and application of the additional face candidates' detection means like motion and color segmentation.

6. REFERENCES

- [1] M. H. Yang. Recent Advances in Face Detection. IEEE ICPR 2004 Tutorial, Cambridge, United Kingdom, 2004, 93 p.
- [2] H. Rowley, S. Baluja, and T. Kanade. Neural network-based face detection // IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.20, 1998, pp.22–38.
- [3] P. Viola, M. Jones. Robust Real-Time Face Detection // International Journal of Computer Vision, 57 (2), 2004, pp. 137–154.
- [4] C. Garcia, M. Delakis. A Neural Architecture for Fast and Robust Face Detection // IEEE-IAPR International Conference on Pattern Recognition (ICPR2002), Québec City, 2002, pp. 40-43.
- [5] Y. LeCun, L. Bottou, Y. Bengio and etc. Gradient-Based Learning Applied to Document Recognition // Intelligent Signal Processing, IEEE Press, 2001, pp. 306-351.
- [6] I. Paliy, A. Sachenko, Y. Kurylyak, V. Turchenko, V. Kapura. Face Detection Using Combined Cascade of Classifiers for Video Surveillance // Scientific journal of the Brest State Technical University: physics, mathematics and informatics series, 2007, pp. 5-8 (in Russian).
- [7] S. Z. Li, Z. Q. Zhang. FloatBoost Learning and Statistical Face Detection // IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, No. 9, September 2004, pp. 1112-1123.
- [8] R. Feraud, O. Bernier, J. Viallet, M. Collobert. A Fast and Accurate Face Detector Based on Neural Networks // IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no.1, 2001, pp. 42-53
- [9] P. Kudryashov. Human Face Detection Algorithms for Image Analysis and Processing Applications. Abstract of PhD. Thesis, Volgograd, 2007, 23 p. (in Russian).
- [10] V. Golovko. Neurointelligence: theory and application. Brest, vol.1, 1999.