ПЕРЕНЕСЕННЯ АДАПТИВНОЇ АНОМІНІЗАЦІЇ У КОНТЕКСТІ КЛАСИФІКАЦІЇ ЗОБРАЖЕНЬ

О.В. АБАШКІН

Дана робота присвячується явищу таргетованого перенесення змагальних атак та перетину цього явища з задачею адаптивної анонімізації в контексті класифікації зображень.

Методи, засновані на глибокому навчанні, стали стандартом для різних задач комп'ютерного зору. Тим не менш, вони неодноразово демонстрували свою вразливість до різних форм порушень вхідних даних, таких як модифікація пікселів, анонімізація областей, які тісно пов'язані з змагальними атаками. У статті[1] запропонований градієнтний алгоритм що усуває вразливість класифікатору до зміни у вхідних даних і водночас приховує чутливу інформацію.

Дане дослідження, має на меті перевірити подібність переносимості адаптивної анонімізації та перенесимості змагальних атак, оскільки ці процеси є подібними у частині модифікації вхідного зображення, а, також пропонує покращення базового градієнтного алгоритму, через збільшення кількості моделей в процесі адаптації.

Вище згаданий алгоритм може бути показаний як оптимізаційна проблема:

$$\widetilde{x}^*_{blur} = \text{argmin}_{\widetilde{x}_{blur} \in X^I_{x_{or}}} d(f_1(\widetilde{x}_{blur}), f_2(\widetilde{x}_{blur}), f_1(x_{or}), f_2(x_{or})), \qquad (1)$$

де I - область анонімізації, а $X^I_{x_{or}} = \left\{ \; x \in X | x[i] = x_{or}[i], i \notin I \; \right\}$ множина анонімізованих зображень. Нехай $f_i : X \to [0,1]^N$ класифікатор на основі нейронної мережі, який повертає вектор оцінок, пов'язаних із задачею класифікації на N класів.

Проміжні результати показали, що повна переносимість анонімізації не може бути досягнута базовим алгоритмом. Результати розподілені наступним чином – Відсоток вдало перенесених прикладів: MobileNetV2-050 - 32%, Resnet152 - 71%, Resnet101 - 69%, VGG16 - 29%, VGG19-BN - 36%. Покращення алгоритму має на меті досягнути кращих результатів шляхом розширення кількості моделей під час процесу адаптації.

Література

[1] Shvai N., Carmona A. L., Nakib A. *Adaptive Image Anonymization in the Context of Image Classification with Neural Networks.* //University Paris Est Creteil, Laboratoire LISSI, 2023. — 5074-5083 с.

Національний університет «Києво-Могилянська академія», Київ, Україна
Email address: o.abashkin@ukma.edu.ua