Maskowanie cyfrowe zobrazowań satelitarnych

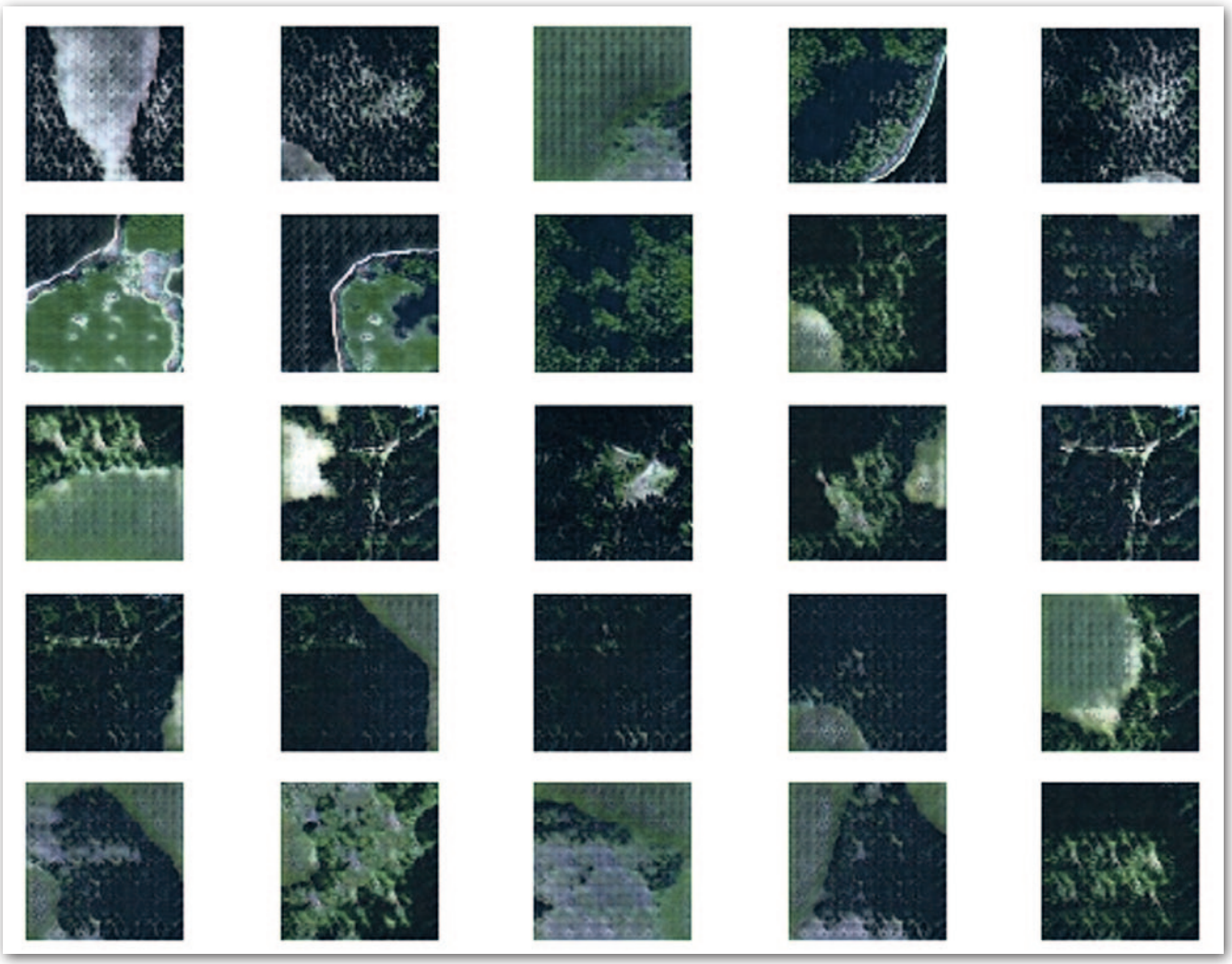
Sztuka ukrywania

Kiedyś na zdjęciach lotniczych i satelitarnych obiekty tajne zaczerniano. Dziś ten sam efekt można osiągnąć, wykorzystując algorytmy sztucznej inteligencji, które pozwalają na ukrycie wrażliwych informacji przed analizą wizualną czy zmylenie klasyfikatorów stosujących głębokie sieci neuronowe.

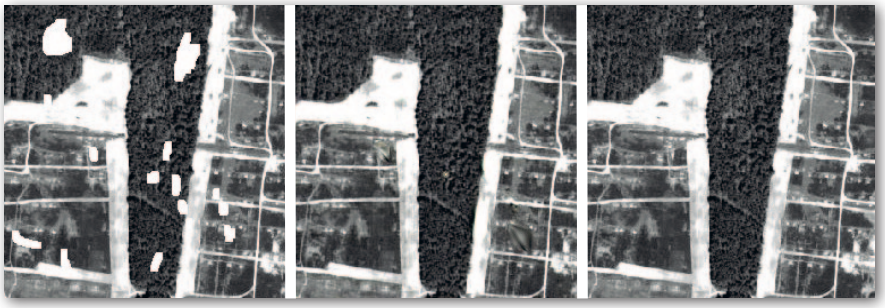
Kinga Reda

Różnego rodzaju algorytmy towarzyszą nam na każdym kroku. Są w naszych komputerach i telefonach, ale też zarządzają produkcją czy decydują, kiedy zapali się zielone światło na skrzyżowaniu. Ich rozwój nie byłby możliwy, gdyby nie dynamiczny wzrost ilości pozyskiwanych danych oraz rosnąca moc obliczeniowa komputerów, które przyczyniły się do znacznego ulepszenia istniejących rozwiązań za pomocą uczenia maszynowego.

Klasyczne algorytmy pracują w następujący sposób: komputery pobierają do obliczeń dane wejściowe, a następnie dostarczają wynik swoich działań w postaci danych wyjściowych. Natomiast w uczeniu maszynowym (czy też sieciach neuronowych, będących podkategorią uczenia maszynowego) algorytm na wejściu pobiera zarówno dane wejściowe, jak i wyjściowe (czyli np. w przypadku maskowania pobiera zarówno zdjęcie z obszarem ukrytym, jak i oryginalne). Na tej podsta-



Rys. 1. Obrazy RGB stworzone za pomocą generatora



Rys. 2. Wynik działania GMCNN dla obrazów panchromatycznych. Z lewej – obraz z utraconymi fragmentami; na środku – obraz po zastosowaniu GMCNN; z prawej – obraz oryginalny

wie podczas szkolenia sieci dobierane są wagi. Dzięki nim po zakończeniu treningu możliwe jest wczytanie zamaskowanego obrazu i odtworzenie (stworzenie) jego ukrytych fragmentów. W efekcie powstaje nowa (zazwyczaj lepsza) wersja algorytmu. Dzięki temu m.in. komputery mogą dzisiaj klasyfikować i wykrywać obiekty na zdjęciach czy też generować nowe obrazy.

W ramach rozpoznania obrazowego pozyskiwanych jest coraz więcej cyfrowych zobrazowań satelitarnych. Mając do dyspozycji tak dużą ilość danych oraz potężne algorytmy, znacznie łatwiej jest przetwarzać te dane, co może prowadzić do tworzenia nowych, fałszywych informacji. Konieczne jest zatem zabezpieczenie danych obrazowych, co utrudni działania „wrogich/obcych" algorytmów wykorzystujących metody sztucznej inteligencji np. w celu ich klasyfikacji. Ponadto – uwzględniając rosnącą liczbę publikowanych zobrazowań satelitarnych oraz zdjęć lotniczych w otwartych źródłach informacji – ważne jest stworzenie algorytmów pozwalających na automatyczne maskowanie elementów oraz obiektów, których lokalizacja jest tajna. Chodzi tu o tereny zamknięte niezbędne dla obronności państwa, o których mowa m.in. w art. 10 *Prawa geodezyjnego i kartograficznego*.

Obraz prawdziwy i wygenerowany

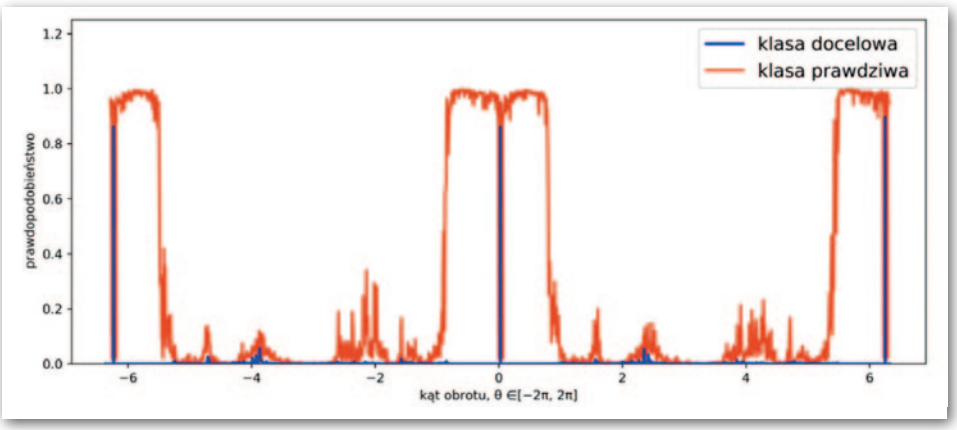
A zatem do stworzenia nowych obrazów można posłużyć się generatywnymi sieciami antagonistycznymi. Opierają się one na teoretycznym scenariuszu gry – wyniki pracy algorytmów rywalizują w modelu dyskryminującym z obrazami prawdziwymi. Przyjmuje się, że sieć generatora została dobrze wyszkolona, gdy sieć dyskryminatora nie jest w stanie odróżnić obrazu prawdziwego od wygenerowanego. Biorąc pod uwagę tę zależność, w ramach pracy magisterskiej przygotowano algorytm pozwalający na wygenerowanie zupełnie nowych, fałszywych obrazów. Wykorzystano dwie bazy obrazów treningowych, które powstały na podstawie udostępnionego przez Ośrodek Rozpoznania Obrazowego w Białobrzegach zobrazowania satelitarnego przedstawiającego Centrum Szkolenia Wojsk Lądowych Drawsko. Pierwsza z nich składała się z 10 061 zdjęć panchromatycznych, druga zaś z 4317 obrazów RGB 128 x 128 pikseli (ich rozmiar został dobrany tak, aby proces treningu jednego modelu nie trwał dłużej niż 100 godzin).

Ostatnim etapem przygotowania bazy danych jest wykonanie augmentacji. Jest to proces umożliwiający zwiększenie liczby i różnorodności danych. Nie polega on na tworzeniu zupełnie nowych obrazów, lecz przekształceniu już istniejących, co w efekcie pozwala na zwielokrotnienie danych. Na podstawie przygotowanych baz w ciągu ponad 170 godzin wyszkolono dwa modele generatora składające się z jedynie pięciu warstw splotowych umożliwiających generowanie nowych, fałszywych obrazów.

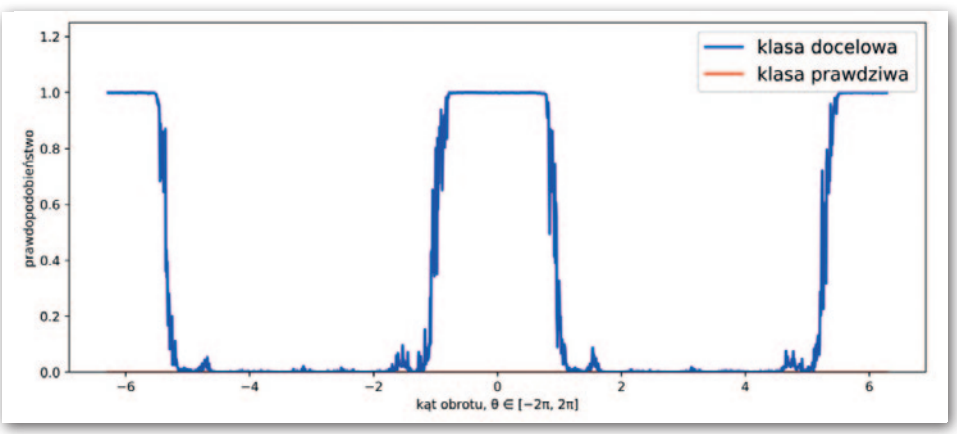
Maskowanie wybranych części obrazu

Innym sposobem na wykonanie maskowania na zobrazowaniach cyfrowych jest ukrycie jedynie wybranych fragmentów obrazu. Można w tym celu wykorzystać generatywne multikolumnowe splotowe sieci neuronowe (GMCNN). W odróżnieniu od modelu odpowiedzialnego za generowanie nowych obrazów w tym przypadku sieć generatora składa się z trzech kolumn, w których do wypełnienia brakujących fragmentów obrazu wykorzystywana jest różna wielkość jądra filtrów. Dodatkowo w GMCNN zastosowany został dyskryminator lokalny (oceniający wypełnienie zamaskowanych fragmentów) oraz dyskryminator globalny (oceniający całość obrazu).

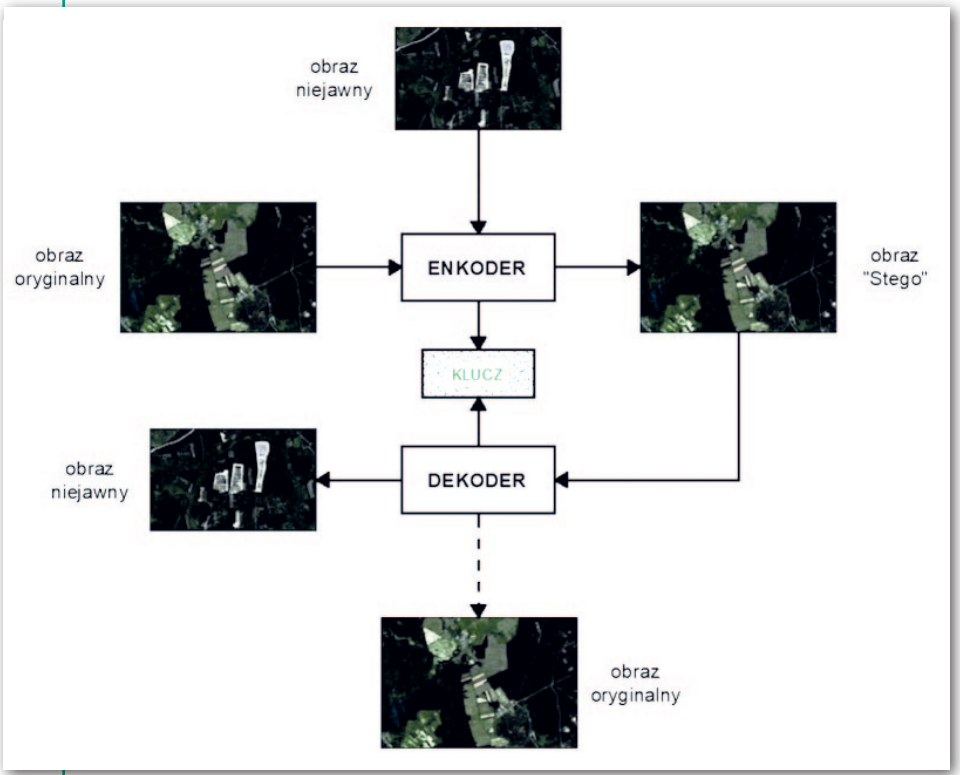
Spośród parametrów definiujących przebieg treningu sieci szczególną uwagę należy zwrócić na dwa: wielkość wsadu oraz szybkość uczenia. Pierwszy z nich mówi o liczbie obrazów branych pod uwagę podczas pojedynczej aktualizacji wag.



Rys. 3. Wpływ kąta obrotu obrazu na prawdopodobieństwo klasyfikacji obrazu przeciwstawnego wykonanego metodą Jacobian Saliency Map Attack za pomocą sieci Inception V3



Rys. 4. Wpływ kąta obrotu obrazu na prawdopodobieństwo klasyfikacji obrazu przeciwstawnego wykonanego metodą Synthesizing Robust za pomocą sieci Inception V3



Rys. 5. Steganografia – zasada działania

Mała liczba obrazów powoduje częstszą poprawę wag modelu, dzięki czemu będzie się on uczył szybciej, ale niekoniecznie lepiej. W tym przypadku dużą rolę odgrywają obrazy znajdujące się w bazie danych treningowych, które znacznie różnią się od pozostałych bądź zawierają błędy – wagi aktualizowane są na podstawie złych wartości. Przy większej wartości tego parametru wagi są aktualizowane na podstawie średniej ze wszystkich obrazów pojedynczego wsadu, dzięki czemu gradienty są dokładniejsze, a wpływ obrazów błędnych – mniejszy.

Drugi parametr, czyli szybkość uczenia, określa wartość, o jaką będą zmieniane wagi modelu sieci w kierunku obliczanego gradientu – im większe tempo, tym szybciej zlokalizowane zostanie minimum badanego problemu. Jednocześnie może się to wiązać z pominięciem ekstremum badanej funkcji, w wyniku czego otrzymane wyniki będą błędne.

W przypadku obrazów panchromatycznych zadowalające wyniki można otrzymać już po 189 tysiącach kroków treningu sieci (ok. 57 godzin). Natomiast w przypadku obrazów RGB szkolenie modelu jest znacznie trudniejsze. W rozwiązaniu tego problemu pomocne jest wykorzystanie połączeń szczątkowych, które pozwalają na przekazanie danych wyjściowych wcześniejszej warstwy do danych wyjściowych warstwy późniejszej. Dzięki temu ograniczone jest powstanie wąskich gardeł reprezentacji oraz problemu zaniku gradientu.

Obrazy przeciwstawne

Współczesne algorytmy klasyfikacji obrazów (takie jak AlexNET, VGG-16 czy Inception-V3) lepiej radzą sobie z tym zadaniem niż człowiek. Co więcej, do wykonania klasyfikacji jednego zdjęcia potrzebują ok. 10 ms (znacznie mniej niż doświadczony analityk obrazu). Jednak istnieją sposoby pozwalające na zabezpieczenie obrazów cyfrowych przed działaniem klasyfikatorów zbudowanych na bazie sieci neuronowej w taki sposób, aby przyporządkowywały zdjęcia do innych, docelowych klas. Należą do nich metody maskowania za pomocą obrazów przeciwstawnych. W wyniku ich działania z wysokim prawdopodobieństwem obiekt na obrazie zostanie sklasyfikowany jako inny.

Jedna z tych metod – Jacobian Saliency Map Attack – działa poprzez nasycenie kilku pikseli w danym obrazie do ich maksymalnych lub minimalnych wartości, dzięki czemu model zaklasyfikuje obraz do błędnej klasy. Wadą JSMA jest brak odporności na występowanie szumów, przesunięć, obrotów czy transformacji. Z problemem tym radzi sobie metoda Synthesizing Robust, a to dzięki wprowadzeniu algorytmu Expectation Over Transformation (EOT), którego zadaniem jest modelowanie zaburzeń podczas procedury optymalizacji. Istotną zaletą obrazów przeciwstawnych, które powstały za pomocą powyższych metod, jest to, że nie są one odróżnialne od swoich oryginalnych wersji. Co więcej, wykrycie manipulacji przy użyciu procedur optymalizacji czy też nasycenia pojedynczych pikseli do wartości ekstremalnych jest czynnością bardzo trudną.

Wszystko jest kwestią klucza

Steganografia jest sposobem kodowania wiadomości i po raz pierwszy została wykorzystana przez Trithemiusa już ok. 1500 r. To metoda ukrywania tajnych, wrażliwych danych poprzez osadzenie ich np. w pliku tekstowym, audio, wideo lub obrazie. Nad kryptografią ma tę przewagę, że wizualnie modyfikacja nie jest widoczna, przez co pliki te nie budzą zainteresowania. Obarczona jest jednak znaczącą wadą – jest łatwiejsza do rozszyfrowania. W przypadku steganografii wiadomość (obraz) ukrywana jest dzięki zmianie wartości niektórych pikseli wybranych przez algorytm szyfrowania. Aby ją rozszyfrować, odbiorca musi wiedzieć, w których pikselach została ukryta.

Steganografia pozwala na ukrycie wrażliwego obrazu w innym, lecz należy pamiętać, że po odkodowaniu widocznemu pogorszeniu ulegnie zarówno obraz z informacją niejawną, jak i obraz, który ją przenosi. Co więcej, należy zwrócić uwagę na metodę kodowania. Może bowiem zdarzyć się, że widoczne będzie prześwitywanie obrazu ukrywanego, szczególnie gdy składa się on z jasnych pikseli.

Człowiek a algorytm

Dzięki rozwojowi technologii algorytmy opierające się na uczeniu maszynowym towarzyszą nam niemal na każdym kroku. Z upływem czasu zdobywają nowe umiejętności – mogą klasyfikować obrazy, wykrywać obiekty, generować nowe obrazy, jak również wypełniać brakujące fragmenty obrazu. Choć proces ich przygotowania i szkolenia wymaga sporej ilości czasu, w efekcie znacznie przewyższają umiejętności wykwalifikowanego człowieka.

Kinga Reda
Zakład Teledetekcji, Fotogrametrii
i Rozpoznania Obrazowego,
Wojskowa Akademia Techniczna

Artykuł powstał na podstawie pracy magisterskiej pt. „Maskowanie cyfrowe zobrazowań satelitarnych" napisanej pod kierunkiem płk. prof. Michała Kędzierskiego i obronionej w 2020 r. Praca została wyróżniona I nagrodą w Konkursie Ministra Obrony Narodowej na najlepszą pracę inżynierską, magisterską i rozprawę doktorską z zakresu technologii, technik i inżynierii kosmicznej i satelitarnej oraz systemów autonomicznych, mających potencjał zastosowania w obszarze obronności lub bezpieczeństwa państwa