BIAŁA KSIĘGA

Sztuczna inteligencja w analizie wideo

Uwagi dotyczące analizy opartej na uczeniu maszynowym i głębokim uczeniu

Marzec 2021

AXIS COMMUNICATIONS

Spis treści

1 Podsumowanie

W branży dozoru wizyjnego jednym z najgoręcej dyskutowanych tematów jest analiza wideo oparta na sztucznej inteligencji. Niektóre aplikacje mogą znacznie przyspieszyć analizę danych i zautomatyzować powtarzalne czynności. Jednak obecne rozwiązania AI nie są w stanie zastąpić doświadczenia i umiejętności decyzyjnych człowieka pełniącego funkcję operatora. W rzeczywistości atrakcyjne jest odpowiednie połączenie, czyli wykorzystanie zalet rozwiązań AI w celu polepszenia jakości pracy i podniesienia efektywności operatorów.

Sztuczna inteligencja to pojęcie łączące algorytmy uczenia maszynowego i głębokiego uczenia. Oba rodzaje algorytmów automatycznie tworzą model matematyczny przy użyciu dużej ilości danych próbnych (tzw. *danych szkoleniowych)*, aby zyskać możliwość obliczania wyników bez odrębnego programowania do tego celu. Algorytm AI powstaje w wyniku procesu o charakterze iteracyjnym, w ramach którego kroki obejmujące zebranie danych szkoleniowych, ich oznakowanie, użycie tych danych do przeszkolenia algorytmu i przetestowanie przeszkolonego algorytmu są powtarzane do czasu osiągnięcia oczekiwanego poziomu jakości. Następnie algorytm jest gotowy do użycia w aplikacji analitycznej, którą można kupić i wdrożyć w obiekcie objętym dozorem. Na tym etapie proces szkolenia jest zakończony i aplikacja nie uczy się już niczego nowego.

W przypadku aplikacji do analiz wideo opartych na sztucznej inteligencji typowym zadaniem jest wizualne wykrywanie ludzi i pojazdów w strumieniu wideo oraz rozróżnianie między tymi rodzajami obiektów. Algorytm *uczenia maszynowego* nauczył się charakterystycznego dla tych obiektów połączenia cech wizualnych. Algorytm *głębokiego uczenia* jest bardziej zaawansowany i po odpowiednim przeszkoleniu może wykrywać znacznie bardzie złożone obiekty. Wymaga też jednak znacznie większego nakładu pracy na etapach programowania i szkolenia oraz znacznie większych zasobów obliczeniowych do obsługi gotowej aplikacji. Dlatego w przypadku jasno określonych potrzeb związanych z dozorem warto się zastanowić, czy nie wystarczy specjalnie zoptymalizowana aplikacja z zakresu uczenia maszynowego.

Ciągły rozwój algorytmów i rosnąca moc obliczeniowa kamer umożliwiły prowadzenie zaawansowanych analiz wideo opartych na sztucznej inteligencji bezpośrednio w kamerze (analizy brzegowe) zamiast wykonywania niezbędnych obliczeń na serwerze (analizy serwerowe). Polepsza to działanie aplikacji w czasie rzeczywistym, ponieważ mają one natychmiastowy dostęp do nieskompresowanego materiału wizyjnego. Dzięki umieszczeniu w kamerach specjalnych akceleratorów sprzętowych, takich jak MLPU (machine learning processing unit — jednostka przetwarzania uczenia maszynowego) i DLPU (deep learning processing unit — jednostka przetwarzania głębokiego uczenia), analizy brzegowe można prowadzić przy mniejszym zużyciu energii w porównaniu z używaniem wyłącznie procesora głównego (CPU) lub graficznego (GPU).

Zanim zostanie zainstalowana aplikacja do analiz wideo wykorzystująca sztuczną inteligencję, należy szczegółowo przeanalizować zalecenia producenta oparte na znanych warunkach wstępnych i ograniczeniach oraz zadbać o ich przestrzeganie. Każdy system dozoru jest wyjątkowy i dlatego należy ocenić wydajność aplikacji w każdym obiekcie. Jeśli się okaże, że jakość nie spełnia oczekiwań, możliwych powodów należy poszukać w sposób całościowy, nie skupiając się jedynie na samej aplikacji analitycznej. Wydajność analizy wideo zależy od wielu czynników, takich jak elementy sprzętowe kamery, jej konfiguracja, jakość materiału wizyjnego, dynamika sceny i oświetlenie. Poznanie wpływu tych czynników i ich odpowiednia optymalizacja pozwalają w wielu przypadkach podnieść wydajność analiz wideo w danej instalacji.

W sytuacji, gdy rośnie popularność sztucznej inteligencji w systemach dozoru, dostrzeganiu zalet związanych z efektywnością operacyjną i nowymi zastosowaniami musi towarzyszyć wyważona dyskusja na temat tego, kiedy i gdzie warto stosować tę technologię.

2 Wprowadzenie

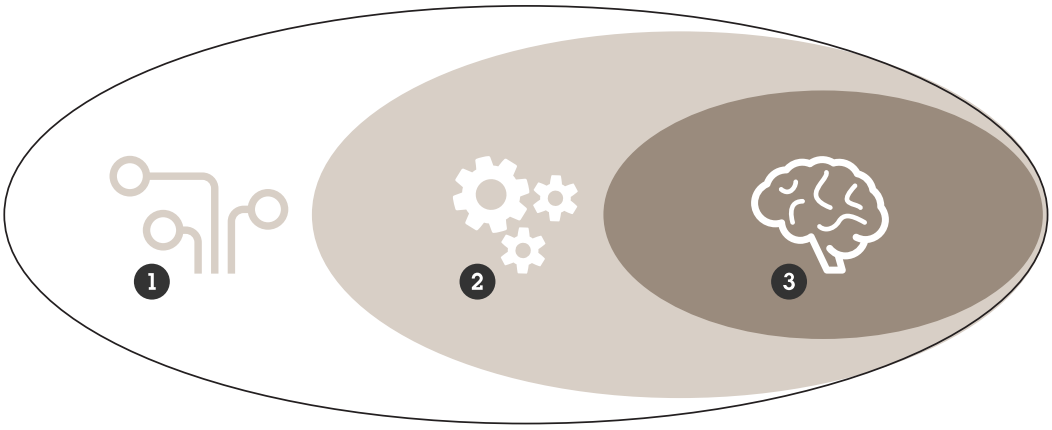
Sztuczna inteligencja (artificial intelligence, AI) to dziedzina rozwijana i dyskutowana, odkąd wynaleziono komputery. I chociaż wciąż czekamy na naprawdę rewolucyjne wcielenia sztucznej inteligencji, technologie AI są obecnie często używane do wykonywania jasno określonych zadań np. w aplikacjach do rozpoznawania głosu, wyszukiwarkach i asystentach wirtualnych. Sztuczna inteligencja jest też coraz częściej stosowana w ochronie zdrowia, gdzie stanowi wartościowe narzędzie np. w diagnostyce rentgenowskiej i analizie skanów siatkówki.

W branży dozoru wizyjnego jednym z najgoręcej dyskutowanych tematów jest analiza wideo oparta na sztucznej inteligencji, z którą wiąże się duże nadzieje. Na rynku są dostępne aplikacje, które przy użyciu algorytmów AI z powodzeniem przyspieszają analizę danych i automatyzują powtarzalne czynności. Jeśli jednak chodzi o szerszy kontekst systemów dozoru, obecnie i w najbliższej przyszłości sztuczną inteligencję należy traktować jako jeden z kilku elementów w procesie tworzenia efektywnych rozwiązań.

W tym dokumencie przedstawiono podstawowe informacje techniczne na temat algorytmów uczenia maszynowego i głębokiego uczenia, a także możliwości ich rozwijania i stosowania w obszarze analizy wideo. Obejmuje to krótkie omówienie sprzętowych akceleratorów sztucznej inteligencji oraz zalet i wad uruchamiania narzędzi analitycznych opartych na technologiach AI na brzegu sieci w porównaniu z serwerem. W dokumencie przedstawiono także możliwe sposoby optymalizacji warunków wstępnych, które wpływają na wydajność analiz wideo opartych na sztucznej inteligencji, biorąc pod uwagę szeroki zakres czynników.

3 Sztuczna inteligencja, uczenie maszynowe i głębokie uczenie

Sztuczna inteligencja (AI) to szerokie pojęcie odnoszące się do maszyn, które potrafią wykonywać złożone zadania, wykazując pozornie inteligentne cechy. Głębokie uczenie i uczenie maszynowe to obszary wchodzące w skład technologii AI.



1 Sztuczna inteligencja
2 Uczenie maszynowe
3 Głębokie uczenie

3.1 Uczenie maszynowe

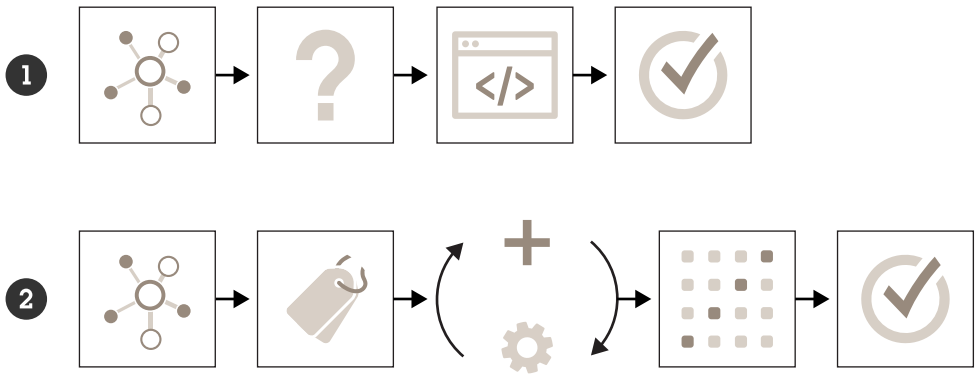
Uczenie maszynowe to obszar sztucznej inteligencji, w ramach którego przy użyciu statystycznych algorytmów uczenia tworzy się systemy zdolne do automatycznego uczenia i doskonalenia się przez szkolenie bez wykonywanych wprost czynności programistycznych.

W tej sekcji przestrzegamy rozróżnienia między tradycyjnym programowaniem i uczeniem maszynowym w kontekście *widzenia komputerowego*, czyli dyscypliny polegającej na doprowadzaniu komputerów do rozumienia zdarzeń zachodzących w obserwowanej scenie przez analizę zdjęć lub nagrań wideo.

W tradycyjnym ujęciu programistycznym widzenie komputerowe opiera się na metodach, które obliczają *cechy* obrazu: przykładem są programy szukające wyraźnych krawędzi i narożników. Cechy te musi ręcznie zdefiniować programista algorytmu, który wie, co jest ważne w danych obrazu. Następnie programista łączy te cechy, aby algorytm mógł określić, co znalazł w obserwowanej scenie.

Algorytmy uczenia maszynowego automatycznie tworzą model matematyczny przy użyciu dużej ilości danych próbnych — tzw. *danych szkoleniowych* — aby zyskać możliwość podejmowania decyzji przez obliczanie wyników bez odrębnego programowania. Cechy nadal są określane ręcznie, ale sposobu ich łączenia uczy się już sam algorytm, który ma styczność z dużą ilością oznakowanych, czyli *adnotowanych*, danych szkoleniowych. W tym dokumencie technikę polegającą na używaniu w wyuczonych połączeniach cech określonych ręcznie nazywamy *klasycznym uczeniem maszynowym*.

Innymi słowy, aby umożliwić uczenie maszynowe, należy przeszkolić komputer w celu uzyskania żądanego programu. W tym układzie człowiek zbiera i adnotuje dane, w czym czasem pomaga wstępna adnotacja wykonywana przez serwer. Wyniki są wprowadzane do systemu i cały proces jest kontynuowany do momentu, aż aplikacja nauczy się wystarczająco dużo, by była w stanie wykryć żądany obiekt, np. określony rodzaj pojazdu. Przeszkolony model staje się programem. Warto zwrócić uwagę, że gdy program zostanie ukończony, system nie uczy się już niczego nowego.



1 *Tradycyjne programowanie:*
Zebrać dane. Zdefiniować kryteria programu. Napisać kod programu (czynność wykonywana przez człowieka). Gotowe.
2 *Uczenie maszynowe:*
Zebrać dane. Oznakować dane. Poddać model iteracyjnemu procesowi szkolenia. Sfinalizowany i przeszkolony model staje się programem. Gotowe.

Z perspektywy tworzenia programu do widzenia komputerowego sztuczna inteligencja ma tę przewagę nad tradycyjnym programowaniem, że umożliwia przetwarzanie ogromnych ilości danych. Komputer może przetworzyć tysiące obrazów bez utraty koncentracji, podczas gdy programista po pewnym czasie się męczy i nie potrafi się dobrze skupić. Dlatego sztuczna inteligencja pozwala stworzyć znacznie

dokładniejszą aplikację. Z drugiej strony im bardziej skomplikowana aplikacja, tym trudniej komputerowi uzyskać żądany rezultat.

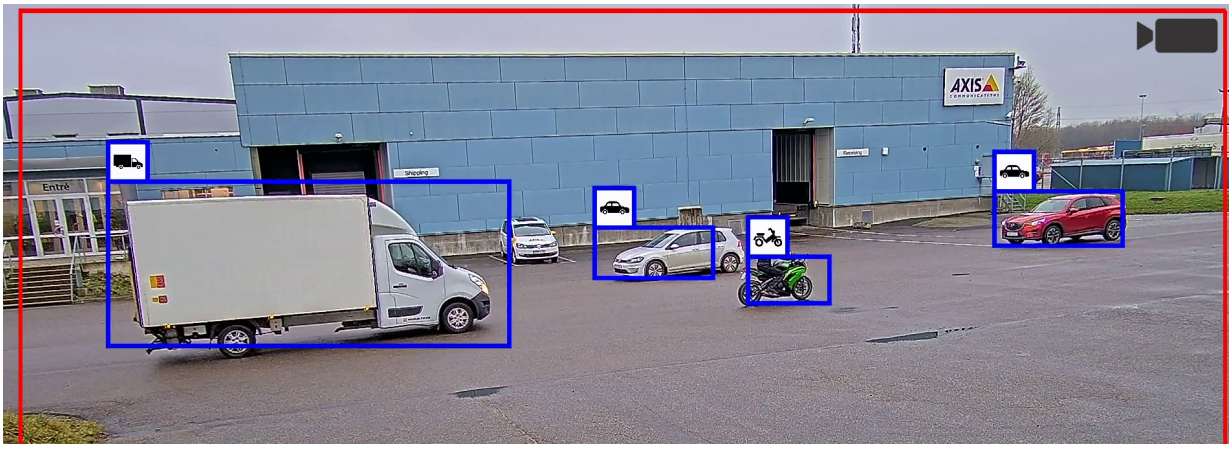
3.2 Głębokie uczenie

Głębokie uczenie jest udoskonaloną odmianą uczenia maszynowego, w której zarówno wyodrębnianie cech, jak i sposób ich łączenia w głębokie struktury reguł w celu uzyskania danych wyjściowych są wyuczane na podstawie danych. Algorytm może automatycznie określić, jakich cech należy szukać w danych szkoleniowych. Potrafi też uczyć się bardzo głębokich struktur obejmujących łańcuchowe połączenia cech.

Główną inspiracją dla algorytmów wykorzystywanych w głębokim uczeniu jest sposób działania neuronów oraz to, jak przy ich użyciu ludzki mózg tworzy wiedzę wyższego poziomu, łącząc sygnały wyjściowe neuronów w ramach głębokiej hierarchii, czyli *sieci*, połączonych łańcuchowo reguł. Mózg to system, w którym również same połączenia są tworzone przez neurony. Zamazuje to rozróżnienie między wyodrębnianiem cech i ich łączeniem, sprawiając, że w pewnym sensie procesy te stają się tożsame. Drogą symulacji tych struktur badacze stworzyli tzw. *sztuczne sieci neuronowe*, które są najbardziej rozpowszechnionym typem algorytmów w głębokim uczeniu. Krótkie omówienie sieci neuronowych jest dostępne w dodatku zawartym w tym dokumencie.

Algorytmy głębokiego uczenia pozwalają konstruować wyrafinowane detektory wizualne oraz automatycznie je szkolić do wykrywania bardzo złożonych obiektów, niezależnie od skali, kąta obrotu i innych zmiennych.

Ta elastyczność wynika z faktu, że systemy głębokiego uczenia mogą się uczyć na podstawie znacznie większej ilości danych — i to danych dużo bardziej zróżnicowanych — niż klasyczne systemy uczenia maszynowego. W większości osiągają znacznie lepsze wyniki od ręcznie zdefiniowanych algorytmów widzenia komputerowego. To sprawia, że głębokie uczenie szczególnie dobrze nadaje się do rozwiązywania złożonych problemów — takich jak klasyfikacja obrazów, przetwarzanie języka i detekcja obiektów — w przypadku których nawet człowiek będący ekspertem nie jest w stanie łatwo sformułować połączenia cech.



Funkcja detekcji obiektów oparta na głębokim uczeniu potrafi klasyfikować złożone obiekty. W tym przykładzie aplikacja analityczna nie tylko wykrywa pojazdy, ale też klasyfikuje je na podstawie rodzaju.

3.3 Klasyczne uczenie maszynowe a głębokie uczenie

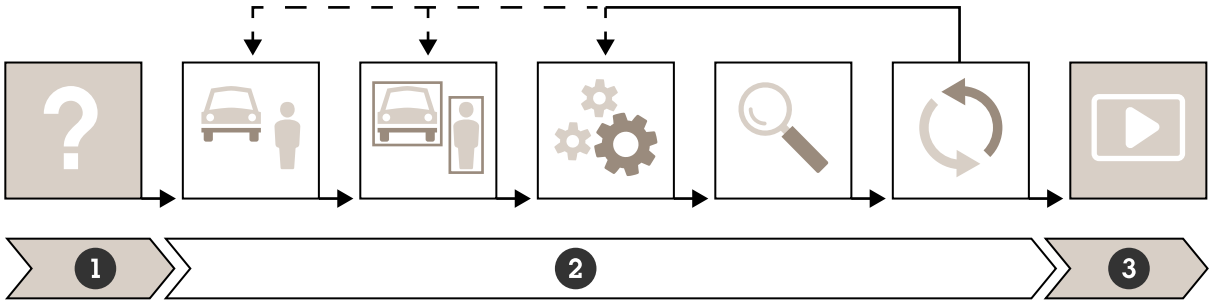
Chociaż algorytmy głębokiego uczenia i klasycznego uczenia maszynowego są do siebie podobne, te pierwsze korzystają ze znacznie obszerniejszego zestawu wyuczonych połączeń cech. Oznacza to, że

narzędzia analityczne oparte na głębokim uczeniu są bardziej elastyczne i przy odpowiednim przeszkoleniu mogą się uczyć wykonywania znacznie bardziej złożonych zadań.

Jeśli jednak chodzi o analizy czysto dozorowe, wystarczający może się okazać specjalny, zoptymalizowany klasyczny algorytm uczenia maszynowego. W odpowiednio określonym zakresie może on zapewnić podobne wyniki co algorytm głębokiego uczenia, wymagając przy tym mniejszej liczby operacji matematycznych, co przekłada się na niższe koszty i mniejsze zużycie energii. Ponadto taki algorytm wymaga znacznie mniej danych szkoleniowych, a to znacznie ogranicza zakres prac programistycznych.

4 Etapy uczenia maszynowego

Podczas prac nad algorytmem uczenia maszynowego wykonywany jest ciąg kroków i iteracji, w zarysie przedstawionych poniżej, które pozwalają stworzyć ostateczną, przeznaczoną do wdrożenia aplikację analityczną. Serce takiej aplikacji stanowi jeden lub kilka algorytmów, na przykład algorytm detekcji obiektów. W przypadku aplikacji opartych na głębokim uczeniu rdzeniem algorytmu jest model głębokiego uczenia.



1 *Przygotowanie: określenie celu aplikacji.*
2 *Szkolenie: zebranie danych szkoleniowych. Dodanie adnotacji do danych. Przeszkolenie modelu. Przetestowanie modelu. Jeśli jakość odbiega od oczekiwań, wcześniejsze kroki są powtarzane w ramach iteracyjnego cyklu doskonalenia.*
3 *Wdrożenie: instalacja i korzystanie z gotowej aplikacji.*

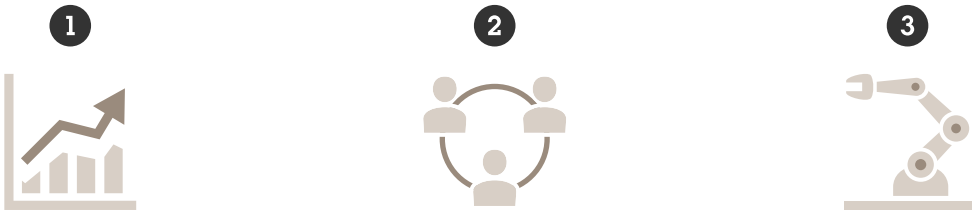
4.1 Zbieranie danych i dodawanie adnotacji

Aby opracować aplikację analityczną opartą na sztucznej inteligencji, należy zebrać dużo danych. W dozorze wizyjnym zazwyczaj są nimi zdjęcia oraz nagrania wideo przedstawiające ludzi i pojazdy lub inne obiekty docelowe. Aby dane stały się rozpoznawalne dla urządzenia lub komputera, trzeba do nich dodać adnotacje, czyli skategoryzować i oznakować odpowiednie obiekty. Opatrywanie danych adnotacjami jest procesem w większości ręcznym i pracochłonnym. Przygotowywane dane muszą obejmować odpowiednio zróżnicowane przykłady adekwatne do kontekstu, w którym będzie używana aplikacja analityczna.

4.2 Szkolenie

W procesie szkolenia, czyli uczenia się, do modelu wprowadzane są dane z adnotacjami, a następnie model jest poddawany iteracyjnym modyfikacjom i ulepszeniom przy użyciu ram szkoleniowych do czasu

osiągnięcia żądanego poziomu jakości. Innymi słowy model jest optymalizowany pod kątem określonego zadania. Szkolenie może być prowadzone przy użyciu jednej z trzech głównych metod.



1 *Uczenie nadzorowane: model uczy się dokładnego przewidywania*
2 *Uczenie nienadzorowane: model uczy się identyfikowania klastrów*
3 *Uczenie ze wzmocnieniem: model uczy się na błędach*

4.2.1 Uczenie nadzorowane

Uczenie nadzorowane to obecnie najczęściej stosowana metoda uczenia maszynowego. Można ją porównać do uczenia się na przykładach. Dane szkoleniowe zawierają jasne adnotacje, co oznacza, że poszczególne dane wejściowe są sparowane z żądanymi wynikami wyjściowymi.

Zazwyczaj uczenie nadzorowane wymaga bardzo dużej ilości danych opatrzonych adnotacjami, a wydajność wyszkolonego algorytmu bezpośrednio zależy od jakości danych szkoleniowych. Z perspektywy jakości najważniejsze jest to, aby użyć zestawu danych reprezentującego wszystkie możliwe dane wejściowe z rzeczywistego wdrożenia. W przypadku detekcji obiektów programista musi zadbać o przeszkolenie algorytmu z uwzględnieniem mocno zróżnicowanych obrazów, obejmujących różne wystąpienia obiektów, orientacje i skale, warunki oświetleniowe, tła oraz czynniki zakłócające. Dane szkoleniowe muszą być reprezentatywne dla planowanego zastosowania — tylko pod tym warunkiem gotowa aplikacja analityczna będzie zapewniać dokładność przewidywania podczas przetwarzania nowych danych, z którymi nie zetknęła się w fazie szkolenia.

4.2.2 Uczenie nienadzorowane

Uczenie nienadzorowane polega na wykorzystaniu algorytmów do analizy i grupowania nieoznakowanych zestawów danych. Ta metoda szkolenia nie jest rozpowszechniona w branży dozoru, ponieważ model wymaga wielu czynności kalibracyjnych i testów, a mimo to jakość bywa nieprzewidywalna.

Zestawy danych muszą być adekwatne dla aplikacji analitycznej, ale nie muszą być jasno oznakowane. Nie ma etapu ręcznego dodawania adnotacji, ale liczbę zdjęć lub nagrań wideo na potrzeby szkolenia należy znacznie zwiększyć: o kilka rzędów wielkości. Na etapie szkolenia model — wspierany przez ramy szkoleniowe — identyfikuje wspólne cechy w zestawach danych. Dzięki temu w fazie wdrożenia model może grupować dane zgodnie z wyuczonymi wzorcami, ale także wykrywać nieprawidłowości niepasujące do żadnej z wyuczonych grup.

4.2.3 Uczenie ze wzmocnieniem

Uczenia ze wzmocnieniem używa się np. w robotyce, automatyce przemysłowej i planowaniu strategii przedsiębiorstw, ale z powodu zapotrzebowania na dużą ilość informacji zwrotnych zastosowania tej metody w dozorze są ograniczone. W uczeniu ze wzmocnieniem chodzi o podejmowanie odpowiednich działań w celu zmaksymalizowania potencjalnej *nagrody* w konkretnej sytuacji — nagroda zwiększa się w miarę dokonywania przez model właściwych wyborów. W przypadku tego algorytmu do szkolenia nie są używane pary dane/oznakowanie, ponieważ sposobem na optymalizację algorytmu jest testowanie jego

decyzji w drodze interakcji z otoczeniem przy jednoczesnym pomiarze nagrody. Celem algorytmu jest nauczenie się zasady podejmowania działań, która przyczynia się do maksymalizacji nagrody.

4.3 Testowanie

Przeszkolony model wymaga szczegółowego przetestowania. Ten etap zazwyczaj obejmuje część zautomatyzowaną, którą uzupełniają obszerne testy wykonywane w rzeczywistych sytuacjach.

W ramach części zautomatyzowanej ustala się parametry wydajności aplikacji przy użyciu nowych zestawów danych, z którymi model nie zetknął się podczas szkolenia. Jeśli parametry te odbiegają od oczekiwań, proces rozpoczyna się od nowa: trzeba zebrać nowe dane szkoleniowe, wprowadzić lub ulepszyć adnotacje i ponownie przeszkolić model.

Po osiągnięciu żądanego poziomu jakości rozpoczyna się test eksploatacyjny. W jego ramach aplikację uruchamia się w rzeczywistych scenariuszach. Liczba scenariuszy i ich wariantów zależy od zakresu aplikacji. Im węższy zakres, tym mniej wariantów należy przetestować. Im szerszy zakres, tym więcej potrzeba testów.

Wyniki ponownie są porównywane i oceniane. Również ten krok może doprowadzić do ponownego rozpoczęcia procesu. Kolejnym możliwym wynikiem może być zdefiniowanie warunków wstępnych, czyli znanego scenariusza, w którym nie zaleca się korzystania z aplikacji albo zalecenie ma tylko charakter częściowy.

4.4 Wdrożenie

Faza wdrożenia jest też nazywana fazą wnioskowania lub przewidywania. *Wnioskowanie* lub *przewidywanie* to proces wykonywania przeszkolonego modelu uczenia maszynowego. Algorytm używa tego, czego się nauczył w fazie szkolenia, aby wygenerować żądane dane wyjściowe. W kontekście analiz w systemach dozoru faza wnioskowania to sytuacja, gdy aplikacja działa w systemie dozoru monitorującym rzeczywiste sceny.

Aby algorytm uczenia maszynowego mógł działać w czasie rzeczywistym na dźwiękowych lub wizyjnych danych wejściowych, zazwyczaj konieczna jest specjalna akceleracja sprzętowa.

5 Analiza brzegowa

Dawniej wysokowydajne analizy wideo były wykonywane na serwerze, ponieważ związane z nimi potrzeby dotyczące mocy i chłodzenia przekraczały możliwości kamer. Jednak w ostatnich latach rozwój algorytmów i rosnąca moc obliczeniowa urządzeń brzegowych umożliwiły prowadzenie zaawansowanych analiz wideo opartych na sztucznej inteligencji na brzegu sieci, czyli w samych kamerach.

Brzegowe aplikacje analityczne mają oczywiste zalety: dostęp do nieskompresowanego materiału wizyjnego uzyskują z bardzo małym opóźnieniem, co umożliwia ich działanie w czasie rzeczywistym, a jednocześnie pozwala uniknąć dodatkowych kosztów i złożonych operacji związanych z przenoszeniem danych do chmury na potrzeby obliczeń. Ponadto analizy brzegowe wiążą się z niższymi kosztami sprzętu i wdrożenia, ponieważ system dozoru wymaga mniejszej ilości zasobów serwerowych.

W niektórych aplikacjach korzystne może się okazać połączenie przetwarzania na brzegu sieci i na serwerze, w ramach którego obróbkę wstępną wykonuje kamera, a za dalsze czynności przetwarzania odpowiada serwer. Taki system hybrydowy może ułatwić ekonomiczną rozbudowę aplikacji analitycznych dzięki obsłudze strumieni z kilku kamer.

6 Akceleracja sprzętowa

Chociaż określoną aplikację analityczną często można uruchamiać na kilku rodzajach platform, użycie dedykowanej akceleracji sprzętowej pozwala uzyskać znacznie wyższą wydajność w warunkach ograniczonej mocy. Akceleratory sprzętowe umożliwiają energooszczędne wdrażanie aplikacji analitycznych. W stosownych okolicznościach mogą być uzupełniane przez zasoby obliczeniowe działające na serwerze lub w chmurze.

- **Procesor graficzny (graphics processing unit, GPU)**. Procesory graficzne powstały przede wszystkim z myślą o obróbce grafiki, ale są też używane do przyspieszania mechanizmów sztucznej inteligencji na serwerach i w chmurze. Chociaż procesory graficzne czasem są używane w systemach wbudowanych (brzegowych), z perspektywy energooszczędności nie są optymalnym rozwiązaniem do wnioskowania opartego na uczeniu maszynowym.
- **Jednostka przetwarzania uczenia maszynowego (machine learning processing unit, MLPU)**. Jednostka MLPU może przyspieszyć wnioskowanie w ramach określonych algorytmów klasycznego uczenia maszynowego przeznaczonych do bardzo energooszczędnego wykonywania zadań z zakresu widzenia komputerowego. Układy MLPU są przeznaczone do realizowanej w czasie rzeczywistym detekcji ograniczonej liczby jednocześnie występujących typów obiektów, takich jak ludzie i pojazdy.
- **Jednostka przetwarzania głębokiego uczenia (deep learning processing unit, DLPU)**. Kamery z wbudowaną jednostką DLPU mogą przyspieszyć energooszczędne wnioskowanie oparte na ogólnych algorytmach głębokiego uczenia, co pozwala na bardziej szczegółową klasyfikację obiektów.

7 Sztuczna inteligencja — technologia na wciąż wczesnym etapie rozwoju

Czasem kusi nas, by porównywać potencjał rozwiązań AI z możliwościami człowieka. Operator systemu dozoru wizyjnego jest w stanie zachować pełną czujność tylko przez krótki czas, ale komputer może nieprzerwanie przetwarzać duże ilości danych z ogromną szybkością, w ogóle się przy tym nie męcząc. Ale założenie, że rozwiązania AI zastąpią operatorów, byłoby zasadniczym nieporozumieniem. Naprawdę atrakcyjne jest realistyczne połączenie, czyli wykorzystanie zalet rozwiązań AI w celu polepszenia jakości pracy i zwiększenia efektywności operatorów.

Często wspomina się, że rozwiązania oparte na uczeniu maszynowym i głębokim uczeniu mogą automatycznie się uczyć lub doskonalić dzięki doświadczeniu. Ale dostępne obecnie systemy AI *nie* nabywają automatycznie nowych umiejętności i *nie* pamiętają wystąpienia konkretnych zdarzeń. Doskonalenie wydajności systemu wymaga jego wielokrotnego doszkalania przy użyciu lepszych i dokładniejszych danych podczas sesji uczenia nadzorowanego. Do uczenia nienadzorowanego zazwyczaj potrzebna jest ogromna ilość danych w celu wygenerowania klastrów, w związku z czym to rozwiązanie nie jest stosowane w systemach dozoru wizyjnego. Obecnie używa się go głównie do analizowania obszernych zestawów danych pod kątem nieprawidłowości, na przykład w transakcjach finansowych. Wiele metod promowanych w obszarze dozoru wizyjnego jako „samouczące" opiera się na statystycznej analizie danych, a nie rzeczywistym wielokrotnym szkoleniu modeli głębokiego uczenia.

W zakresie dozoru ludzkie doświadczenie wciąż przewyższa wiele aplikacji do analiz opartych na sztucznej inteligencji. W szczególności dotyczy to wykonywania bardzo ogólnych zadań i sytuacji, w których kluczowe znaczenie ma zrozumienie kontekstu. Odpowiednio przeszkolona aplikacja oparta na uczeniu maszynowym może z powodzeniem wykryć „biegnącą osobę", ale w odróżnieniu od człowieka, który potrafi osadzić tę informację w kontekście, aplikacja nie zrozumie, dlaczego ta osoba biegnie: czy chce złapać autobus czy ucieka przed ścigającym ją policjantem? Mimo obietnic firm, które stosują technologie AI w

aplikacjach analitycznych przeznaczonych do dozoru, aplikacje te wciąż nie potrafią nawet zbliżyć się do człowieka pod względem rozumienia zawartości materiału wizyjnego.

Z tego samego powodu aplikacje analityczne oparte na sztucznej inteligencji mogą też wyzwalać fałszywe alarmy lub nie wyzwalać alarmów wtedy, gdy powinny. Zazwyczaj dzieje się tak w złożonym, bardzo ruchliwym środowisku. Ale to samo może dotyczyć np. osoby niosącej duży przedmiot, gdy zasłonięcie cech człowieka przed aplikacją obniża prawdopodobieństwo prawidłowej klasyfikacji.

Na obecnym etapie narzędzia analityczne oparte na sztucznej inteligencji powinny pełnić funkcję pomocniczą, na przykład z grubsza określając istotność incydentu, a następnie powiadamiając operatora, który podejmuje decyzję dotyczącą reakcji. W takim podejściu sztuczna inteligencja pozwala osiągnąć dużą skalę analiz, a operator zajmuje się oceną incydentów.

8 Uwagi pomagające uzyskać optymalną wydajność analiz

Aby realistycznie sformułować oczekiwania dotyczące jakości pracy aplikacji analitycznej opartej na technologii AI, warto uważnie zapoznać się ze znanymi warunkami wstępnymi i ograniczeniami, które zazwyczaj są wymienione w dokumentacji aplikacji.

Każdy system dozoru jest wyjątkowy i dlatego należy ocenić wydajność aplikacji w każdym obiekcie. Jeśli jakość odbiega od pierwotnych oczekiwań, w poszukiwaniu przyczyn na pewno nie należy koncentrować się wyłącznie na aplikacji. Wszelkie czynności wyjaśniające należy prowadzić w sposób całościowy, ponieważ wydajność aplikacji analitycznej zależy od wielu czynników, które można zoptymalizować, jeśli ma się świadomość ich wpływu. Czynniki te obejmują np. elementy sprzętowe kamery, jakość materiału wizyjnego, dynamikę sceny i oświetlenie, a także konfigurację, położenie i kierunek ustawienia kamery.

8.1 Użyteczność obrazu

Użyteczność obrazu często kojarzy się z wysoką rozdzielczością i dużą światłoczułością kamery. Chociaż trudno kwestionować znaczenie tych czynników, na pewno istnieją inne, które w równym stopniu wpływają na rzeczywistą *użyteczność* zdjęcia lub nagrania wideo. Przykładowo strumień wideo o najlepszej jakości pochodzący z najdroższej kamery może się okazać bezużyteczny, jeśli scena nie będzie dostatecznie oświetlona w nocy albo dojdzie do zmiany kierunku ustawienia kamery lub zerwania połączenia z systemem.

Przed wdrożeniem należy dobrze rozważyć umiejscowienie kamery. Aby analiza wideo działała zgodnie z oczekiwaniami, położenie kamery musi zapewniać wyraźny, niezasłonięty obraz żądanej sceny.

Użyteczność obrazu może także zależeć od rodzaju zastosowań. Materiał wizyjny, który człowiekowi wydaje się wystarczająco dobry, może nie mieć optymalnej jakości z perspektywy aplikacji analitycznej. W analizie wideo wręcz odradza się stosowanie wielu popularnych metod przetwarzania obrazu, które polepszają wygląd materiału wizyjnego na potrzeby odbioru przez człowieka. Przykładowo może to dotyczyć technik redukcji szumów, metod poszerzających zakres dynamiki czy algorytmów automatycznej ekspozycji.

Wiele dzisiejszych kamer wideo jest wyposażonych w zintegrowaną lampę podczerwieni, która umożliwia im pracę w całkowitej ciemności. To duża zaleta, ponieważ taką kamerę można zainstalować w miejscu o trudnych warunkach oświetleniowych, ograniczając potrzebę montażu dodatkowego oświetlenia. Jeśli jednak w danym miejscu zapowiadane są intensywne opady deszczu lub śniegu, stanowczo nie należy polegać na świetle pochodzącym z kamery lub źródła umieszczonego bardzo blisko niej. Zbyt wiele takiego światła po odbiciu od kropli deszczu lub płatków śniegu może trafić z powrotem do kamery, uniemożliwiając

działanie aplikacji analitycznej. Natomiast światło otoczenia daje większe szanse uzyskania jakichś wyników analiz nawet przy niesprzyjającej pogodzie.

8.2 Odległość detekcji

Trudno jest określić maksymalną odległość detekcji zapewnianą przez aplikację analityczną opartą na technologii AI, ponieważ nawet dokładna wartość w metrach podana w arkuszu danych technicznych nie będzie w pełni miarodajna. Odległość detekcji w znacznym stopniu zależy od jakości obrazu, cech sceny, warunków atmosferycznych oraz takich właściwości obiektu jak kolor i jasność. Przykładowo oczywiste jest, że jasny obiekt na ciemnym tle w słoneczny dzień aplikacja wizyjna wykryje ze znacznie większej odległości niż ciemny obiekt przy deszczowej pogodzie.

Odległość detekcji zależy także od szybkości obiektu. Aby zapewnić dokładne wyniki detekcji, aplikacja do analizy wideo musi „widzieć" obiekt przez wystarczająco długi czas. Jego konkretna długość zależy od wydajności przetwarzania (szybkości klatek) danej platformy: im niższa wydajność przetwarzania, tym dłużej obiekt musi być widoczny, aby aplikacja go wykryła. Jeśli czas migawki w kamerze nie jest dobrze dopasowany do szybkości poruszania się obiektu, rozmycie obrazu spowodowane ruchem może dodatkowo pogorszyć skuteczność detekcji.

Szybkie obiekty są łatwiejsze do przeoczenia, jeśli przemieszczają się bliżej kamery. Przykładowo aplikacja może wykryć biegnącą osobę, która znajduje się w dużej odległości od kamery, natomiast osoba przebiegająca bardzo blisko kamery z tą samą szybkością może się znaleźć w polu widzenia przez tak krótki czas, że nie spowoduje wyzwolenia alarmu.

W analizach opartych na detekcji ruchu kolejną trudność stanowią obiekty przemieszczające się bezpośrednio w stronę kamery lub w kierunku przeciwnym. Detekcja jest szczególnie trudna w przypadku obiektów poruszających się z małą szybkością, które w porównaniu z ruchem w poprzek sceny powodują bardzo małe zmiany w obrazie.

Kamery o wyższej rozdzielczości zazwyczaj nie zapewniają większej odległości detekcji. Moc obliczeniowa potrzebna do obsługi algorytmu uczenia maszynowego jest proporcjonalna do wielkości danych wejściowych. Oznacza to, że do analizy materiału z kamery 4K w pełnej rozdzielczości potrzebna jest co najmniej czterokrotnie większa moc obliczeniowa niż w przypadku kamery 1080p. Ze względu na ograniczone możliwości obliczeniowe kamer aplikacje AI bardzo często wykonują analizy przy niższej rozdzielczości niż dostępna w kamerze lub strumieniu.

8.3 Konfiguracja alarmów i nagrywania

Ze względu na różne poziomy stosowanych filtrów narzędzia do analizy obiektów generują bardzo mało fałszywych alarmów. Działają one jednak właściwie tylko w przypadku spełnienia wszystkich wymienionych warunków wstępnych. W przeciwnym razie mogą nie wykryć niektórych ważnych zdarzeń.

Jeśli więc nie ma absolutnej pewności, że wszystkie warunki będą spełnione za każdym razem, zaleca się bardziej ostrożne podejście, zgodnie z którym system należy skonfigurować tak, aby określona klasyfikacja obiektu nie była jedynym czynnikiem wyzwalającym alarm. Spowoduje to zwiększenie liczby fałszywych alarmów, ale także zmniejszy ryzyko przeoczenia czegoś ważnego. W sytuacji, gdy alarmy lub czynniki wyzwalające są przekazywane bezpośrednio do centrum monitorowania alarmów, każdy fałszywy alarm jest bardzo kosztowny. To oczywiste, że potrzebna jest niezawodna klasyfikacja obiektów, pozwalająca odfiltrowywać zbędne alarmy. Jednak rozwiązanie do nagrywania materiału można i należy skonfigurować tak, aby nie opierało się wyłącznie na klasyfikacji obiektów. W przypadku przeoczenia rzeczywistego alarmu pozwala to ocenić na podstawie nagrania, co doprowadziło do przeoczenia alarmu, a następnie udoskonalić całą instalację i konfigurację.

Jeśli klasyfikacja obiektów jest wykonywana na serwerze podczas wyszukiwania incydentów, zaleca się skonfigurowanie systemu do nagrywania ciągłego i niefiltrowanie początkowego nagrania. Nagrywanie ciągłe zużywa dużo miejsca w pamięci masowej, ale w pewnym stopniu rekompensują to nowoczesne algorytmy kompresji, takie jak Zipstream.

8.4 Konserwacja

System dozoru wymaga regularnych czynności konserwacyjnych. Warto przeprowadzać kontrole fizyczne, a nie ograniczać się tylko do przeglądania materiału wizyjnego w oprogramowaniu do zarządzania, ponieważ dzięki kontrolom można odkryć i usunąć ewentualne przedmioty zmniejszające lub zasłaniające pole widzenia. Jest to ważne również w zwykłych instalacjach, obejmujących wyłącznie funkcje nagrywania, natomiast w razie korzystania z aplikacji analitycznych staje się sprawą krytyczną.

W przypadku podstawowej wizyjnej detekcji ruchu typowa przeszkoda, np. pajęcza sieć drgająca na wietrze, może zwiększyć liczbę alarmów, prowadząc do nadmiernego wykorzystania pamięci masowej. Natomiast w przypadku analizy obiektów taka sieć spowodowałaby utworzenie strefy wykluczenia w obszarze detekcji. Jej nici zasłaniałyby obiekty oraz znacznie zmniejszyłyby szanse na detekcję i klasyfikację.

Pajęczyna może ograniczyć pole widzenia kamery dozorowej.

Brud znajdujący się na przedniej soczewce kamery raczej nie spowoduje problemów za dnia. Jednak w warunkach słabego oświetlenia światło padające na zabrudzoną soczewkę z boku, na przykład z reflektorów przejeżdżającego samochodu, może spowodować nieoczekiwane refleksy, które zmniejszą dokładność detekcji.

Czynności dotyczące obserwowanej sceny są równie ważne jak konserwacja kamery. W całym okresie eksploatacji kamery dozorowana scena może ulegać sporym zmianom. Już proste porównanie obrazów „przed" i „po" pozwoli odkryć ewentualne problemy. Jak wyglądała scena w chwili wdrożenia kamery, a jak wygląda obecnie? Czy trzeba skorygować strefę detekcji? Czy należy poprawić pole widzenia kamery, a może przenieść kamerę w inne miejsce?

9 Prywatność i nienaruszalność osobista

Praca w branży bezpieczeństwa i dozoru wymaga równoważenia praw osób fizycznych do prywatności oraz nienaruszalności osobistej z dążeniem do zwiększania bezpieczeństwa przez zapobieganie przestępczości lub umożliwienie prac wyjaśniających. W konkretnej instalacji i konkretnym zastosowaniu wymaga to starannego namysłu etycznego, a także zrozumienia i przestrzegania lokalnych przepisów. Ponadto samo rozwiązanie powinno np. zapewniać cyberbezpieczeństwo i zapobiegać nieumyślnemu dostępowi do materiału wizyjnego. Jednocześnie analizy brzegowe i procesy generowania metadanych do celów statystycznych mogą wzmacniać ochronę prywatności, jeśli do dalszego przetwarzania przekazywane są wyłącznie dane zanonimizowane.

Wraz z coraz powszechniejszym stosowaniem zautomatyzowanych analiz w systemach dozoru trzeba wziąć pod uwagę kilka nowych aspektów. Ponieważ aplikacje analityczne są obarczone ryzykiem fałszywych wyników detekcji, proces decyzyjny powinien obejmować uczestnictwo doświadczonego operatora lub użytkownika. Jest to tzw. zasada udziału człowieka (human in the loop, HITL). Ponadto trzeba mieć świadomość, że na decyzję człowieka może wpłynąć sposób wygenerowania lub przedstawienia alarmu. Brak odpowiedniego przeszkolenia i świadomości dotyczącej sposobu funkcjonowania rozwiązania analitycznego może doprowadzić do błędnych wniosków.

Kolejne obawy mogą wynikać ze sposobu tworzenia algorytmów głębokiego uczenia, dlatego w niektórych rodzajach ich zastosowań wymagana jest duża ostrożność. Jakość tych algorytmów jest ściśle powiązana z zestawami danych szkoleniowych, czyli nagrań wideo i zdjęć. Testy pokazały, że przy braku starannego doboru materiału niektóre systemy AI mogą podczas detekcji wykazywać „skrzywienie" etniczne i płciowe. Stało się to przyczynkiem do otwartej dyskusji i doprowadziło zarówno do wprowadzenia ograniczeń ustawowych, jak i do podjęcia działań mających na celu uwzględnienie takich aspektów w ramach prac nad systemami.

Sztuczna inteligencja jest coraz częściej wykorzystywana w systemach dozoru, dlatego dostrzeganiu zalet związanych z efektywnością operacyjną i potencjalnymi nowymi zastosowaniami musi towarzyszyć wyważona dyskusja na temat tego, kiedy i gdzie warto stosować tę technologię.

10 Dodatek

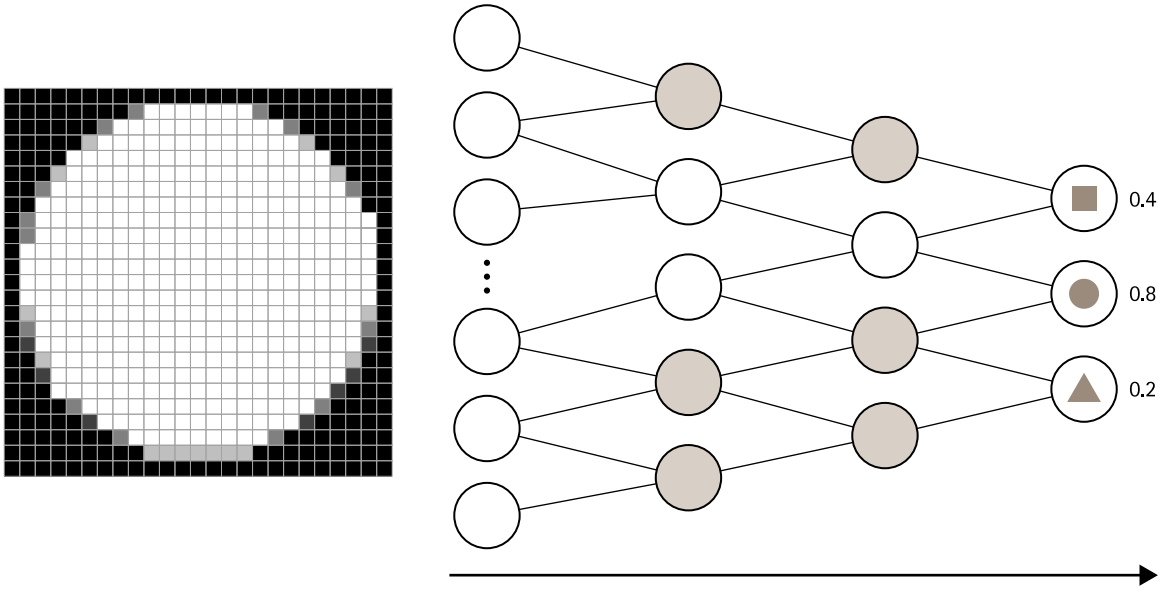
W tym dodatku przedstawiono podstawowe informacje na temat sztucznych sieci neuronowych, które są podstawą głębokiego uczenia.

10.1 Sieci neuronowe

Sieciami neuronowymi nazywa się rodzinę algorytmów, które umożliwiają rozpoznawanie powiązań w zestawach danych przez zastosowanie procesu nieco przypominającego funkcjonowanie mózgu człowieka. Sieć neuronowa składa się z hierarchii wielu warstw wzajemnie połączonych tzw. węzłów lub neuronów. Za pośrednictwem tych połączeń informacje są przekazywane z warstwy wejściowej przez sieć do warstwy wyjściowej.

Aby sieć neuronowa mogła działać, niezbędne jest założenie, że określoną próbkę danych wejściowych można zredukować do skończonego zestawu cech, który stanowi dobrą reprezentację danych wejściowych. Cechy te można następnie połączyć, aby ułatwić klasyfikację danych wejściowych, na przykład opisując zawartość obrazu.

W przykładzie przedstawionym na poniższej ilustracji sieć neuronowa służy do identyfikowania kategorii, do której należy obraz wejściowy. Każdemu pikselowi obrazu odpowiada jeden węzeł wejściowy. Wszystkie węzły wejściowe są połączone z węzłami pierwszej warstwy. Generują one wartości wyjściowe, które są przekazywane jako wartości wejściowe do drugiej warstwy itd. Proces realizowany w każdej warstwie obejmuje także funkcje ważenia, wartości odchylenia i funkcje aktywacji.



Przykład obrazu wejściowego (po lewej) i sieci neuronowej (po prawej). Po dojściu do warstwy wyjściowej sieć określiła prawdopodobieństwo przynależności do poszczególnych kategorii (kwadrat, koło lub trójkąt). O kategorii mającej najwyższą wartość można z najwyższym prawdopodobieństwem powiedzieć, że to do niej należy kształt widoczny na obrazie wejściowym.

Ten proces jest nazywany *propagacją w przód*. W przypadku niezgodności wyniku propagacji w przód parametry sieci są nieznacznie modyfikowane w ramach *propagacji wstecznej*. Ten iteracyjny proces szkolenia prowadzi do stopniowego podnoszenia wydajności sieci.

Po wdrożeniu sieć neuronowa zasadniczo nie pamięta wyników poprzednich przebiegów propagacji w przód. Oznacza to, że sieć nie doskonali się z upływem czasu oraz że może wykrywać tylko te rodzaje obiektów lub rozwiązywać te rodzaje problemów, pod kątem których została przeszkolona.

10.2 Splotowe sieci neuronowe

Splotowe sieci neuronowe (nazywane też konwolucyjnymi) to podtyp sztucznych sieci neuronowych, który szczególnie dobrze sprawdza się w obszarze widzenia komputerowego i jest motorem dynamicznego rozwoju technologii głębokiego uczenia. W przypadku widzenia komputerowego sieć jest szkolona tak, aby automatycznie szukać szczególnych cech obrazu, takich jak krawędzie, narożniki i różnice barwowe, w celu identyfikacji kształtów obiektów na obrazie.

Głównym działaniem umożliwiającym realizację tego celu jest operacja matematyczna nazywana *splotem*. Jest ona bardzo wydajna, ponieważ wartość wyjściowa każdego węzła zależy wyłącznie od ograniczonego obszaru danych wejściowych, wygenerowanego przez poprzednią warstwę, a nie od całego zasobu danych wejściowych. Innymi słowy, w splotowej sieci neuronowej poszczególne węzły są połączone nie ze wszystkimi węzłami poprzedniej warstwy, tylko z ich niewielkim podzbiorem. Sploty są uzupełniane przez inne operacje, które zmniejszają ilość danych przy zachowaniu najbardziej przydatnych informacji. Podobnie jak w zwykłej sztucznej sieci neuronowej, im głębszy stopień zagłębienia w sieć, tym bardziej abstrakcyjne stają się dane.

W fazie szkolenia splotowa sieć neuronowa uczy się optymalnego sposobu stosowania warstw. Chodzi o to, jak operacje splotu powinny łączyć cechy z poprzedniej warstwy, aby dane wyjściowe sieci jak najbardziej odpowiadały adnotacjom zawartym w danych szkoleniowych. Następnie podczas wnioskowania wyszkolona splotowa sieć neuronowa kolejno stosuje warstwy splotów stanowiących rezultat szkolenia.

O firmie Axis Communications

Axis wspiera rozwój inteligentnego oraz bezpiecznego świata poprzez tworzenie rozwiązań sieciowych, które dostarczają wiedzę umożliwiającą poprawę bezpieczeństwa i wdrażanie nowych sposobów prowadzenia działalności. Jako lider rynku sieciowych systemów wizyjnych Axis oferuje produkty i usługi z zakresu dozoru wizyjnego i analiz wideo, kontroli dostępu, systemów domofonowych oraz systemów audio. Axis zatrudnia ponad 3800 pracowników w ponad 50 krajach i współpracuje z partnerami na całym świecie w celu dostarczania swoich rozwiązań klientom. Firma została założona w 1984 roku i ma swoją siedzibę w Lund w Szwecji.

Więcej informacji o Axis można uzyskać odwiedzając stronę internetową firmy *axis.com*