

Object Detection: Evolution of Techniques and Future Directions

Błażej Domagała
Insitute of Informatics
University Of Opole
ul. Oleska 48, 50-204 Opole, Poland *

Abstract

W artykule przeglądowym opartym na "Object Detection in 20 Years: A Survey" [1], omówiono ewolucję technik wykrywania obiektów, poczynawszy od tradycyjnych metod opartych na ręcznie tworzonych cechach, aż do obecnych podejść opartych na uczeniu głębokim. Przedstawiono również przyszłe kierunki badań, takie jak ulepszanie istniejących modeli, rozwój nowych technik adnotacji obrazu, integracja z innymi dziedzinami oraz wykorzystanie nowych technologii. Celem artykułu jest dostarczenie czytelnikom kompleksowej mapy ewolucji wykrywania obiektów oraz wiedzy na temat zarówno przeszłych, jak i przyszłych technik.

1 Wprowadzenie

Wykrywanie obiektów to jedno z najważniejszych i najtrudniejszych zadań w dziedzinie wizji komputerowej, które w ostatnich latach zyskało dużą uwagę. W ciągu ostatnich dwóch dekad zaobserwowaliśmy szybki rozwój technologii wykrywania obiektów i ich głęboki wpływ na całą dziedzinę wizji komputerowej. W niniejszym artykule przeglądowym przedstawimy ewolucję technik wykrywania obiektów oraz kierunki przyszłego rozwoju.

1.1 Cel i zakres artykułu

- Kluczowe detektory w historii: od tradycyjnych metod do technik opartych na uczeniu głębokim
- Zestawy danych, metryki i techniki przyspieszające
- Ewolucja wykrywania wieloskalowego, primingu kontekstowego, wydobycia trudnych negatywów, funkcji straty i tłumienia nie maksymalnego
- Najnowsze osiągnięcia, takie jak wykrywanie bez kotwic, odporność na zmiany skali i rotacji, poprawiona lokalizacja oraz wykrywanie przy użyciu strat segmentacji
- Kierunki przyszłe, takie jak lekkie wykrywanie, wykrywanie end-to-end, wykrywanie małych obiektów, wykrywanie 3D, wykrywanie międzymodalnościowe i wykrywanie w otwartym świecie

1.2 Krótki przegląd historii wykrywania obiektów

Wizja komputerowa, która obejmuje wykrywanie obiektów, zaczęła się rozwijać na początku lat 70. XX wieku, kiedy komputery zaczęły korzystać z określonych algorytmów do przetwarzania obrazów i rozpoznawania podstawowych funkcji. Wcześniejsze podejścia do wykrywania obiektów opierały się na ręcznym projektowaniu cech i klasyfikatorów, takich jak maszyny wektorów nośnych (SVM). Wraz z rozwojem technologii, głębokie sieci neuronowe zaczęły zastępować te ręcznie projektowane cechy, co pozwoliło na automatyczne uczenie się cech i klasyfikacji obiektów.

1.3 Wpływ technik wykrywania obiektów na komputerowe widzenie maszynowe

Techniki wykrywania obiektów stanowią podstawę dla wielu innych zadań komputerowego widzenia maszynowego, takich jak śledzenie obiektów, segmentacja instancji, opis obrazu i inne. Postęp w wykrywaniu obiektów prowadzi także do rozwoju w tych dziedzinach.

*e-mail: 134024@student.uni.opole.pl

2 Ewolucja technik wykrywania obiektów

W pracy "Object Detection in 20 Years: A Survey" [1] autorzy przedstawili przegląd technik wykrywania obiektów, które ewoluowały w ciągu ostatnich 20 lat. Przełomowym momentem w tej dziedzinie było wprowadzenie głębokich sieci neuronowych, które zrewolucjonizowały sposób, w jaki modele AI identyfikują różne obiekty na obrazach.

Wcześniejsze podejścia do wykrywania obiektów opierały się na ręcznym projektowaniu cech i klasyfikatorów, takich jak maszyny wektorów nośnych (SVM). Wraz z rozwojem technologii, głębokie sieci neuronowe zaczęły zastępować te ręcznie projektowane cechy, co pozwoliło na automatyczne uczenie się cech i klasyfikacji obiektów.

3 Klasyczne metody wykrywania obiektów

Przed wprowadzeniem głębokich sieci neuronowych, klasyczne metody wykrywania obiektów opierały się na ręcznie projektowanych cechach i klasyfikatorach. W artykule "Object Detection in 20 Years: A Survey", autorzy przedstawili kilka klasycznych metod wykrywania obiektów, które były stosowane przed erą uczenia głębokiego. Oto niektóre z nich:

3.1 Detektor Viola-Jones

Ten detektor został opracowany w 2001 roku i był jednym z pierwszych skutecznych algorytmów wykrywania obiektów, szczególnie twarzy. Detektor Viola-Jones opiera się na cechach Haar'a, które są prostymi cechami obrazu, takimi jak różnice jasności między prostokątnymi obszarami obrazu. Detektor wykorzystuje kaskadę klasyfikatorów opartych na Adaboost do szybkiego wykrywania obiektów.

3.2 Histogramy zorientowanych gradientów (HOG)

HOG to metoda wykrywania obiektów oparta na analizie gradientów obrazu. HOG dzieli obraz na małe, przesuwne okna, a następnie oblicza histogramy gradientów w każdym oknie. Histogramy te są następnie używane jako cechy do trenowania klasyfikatora, takiego jak maszyna wektorów nośnych (SVM).

3.3 SIFT (Scale-Invariant Feature Transform)

SIFT to algorytm wykrywania i opisywania cech obrazu, który jest niezmienniczy na skalę, rotację i zmiany oświetlenia. SIFT identyfikuje punkty charakterystyczne na obrazie, a następnie oblicza deskryptory oparte na gradientach obrazu wokół tych punktów. Deskryptory te mogą być używane do dopasowywania punktów charakterystycznych między różnymi obrazami, co jest przydatne w wykrywaniu obiektów.

3.4 SURF (Speeded-Up Robust Features)

SURF to usprawniona wersja SIFT, która jest szybsza i bardziej odporna na zmiany oświetlenia. Podobnie jak SIFT, SURF identyfikuje punkty charakterystyczne na obrazie i oblicza deskryptory oparte na gradientach obrazu wokół tych punktów.

Wszystkie te klasyczne metody wykrywania obiektów opierają się na ręcznie projektowanych cechach i klasyfikatorach, takich jak maszyny wektorów nośnych (SVM). Wraz z rozwojem technologii, głębokie sieci neuronowe zaczęły zastępować te ręcznie projektowane cechy, co pozwoliło na automatyczne uczenie się cech i klasyfikacji obiektów.

4 Przełomowe detektory oparte na uczeniu głębokim

Wprowadzenie sieci konwolucyjnych (CNN) w 2012 roku zrewolucjonizowało metody wykrywania obiektów, dzięki zdolności CNN do uczenia się efektywnych, wysokopoziomowych reprezentacji obrazów. W artykule "Object Detection in 20 Years: A Survey", autorzy przedstawili kilka przełomowych technik uczenia głębokiego, które wpłynęły na wykrywanie obiektów. Oto niektóre z nich:

4.1 R-CNN (Region-based Convolutional Neural Networks)

R-CNN to metoda łącząca propozycje regionów z konwolucyjnymi sieciami neuronowymi (CNN) w celu wykrywania obiektów. R-CNN pozwala na lokalizowanie obiektów za pomocą głębokiej sieci i uczenie wysokopoziomowych modeli z niewielką ilością danych anotowanych.

4.2 Fast R-CNN

Fast R-CNN to ulepszona wersja R-CNN, która wprowadza tzw. RoI (Region of Interest) Pooling, co pozwala na szybsze przetwarzanie obrazów i wykrywanie obiektów. Fast R-CNN osiąga wyższą jakość detekcji (mAP) niż R-CNN i SPPnet, jednocześnie będąc szybszym.

4.3 Faster R-CNN

Faster R-CNN to kolejna ewolucja R-CNN, która wprowadza tzw. Region Proposal Network (RPN) do generowania propozycji regionów. Dzięki temu Faster R-CNN osiąga jeszcze lepszą wydajność i dokładność detekcji obiektów.

4.4 SPPNet

SPPNet pozwolił na jednokrotne obliczenie mapy cech dla całego obrazu, zamiast obliczać je indywidualnie dla każdej propozycji obiektu. Dzięki temu SPPNet był ponad 20 razy szybszy od R-CNN, zachowując jednak jego dokładność.

4.5 YOLO (You Only Look Once)

YOLO to szybka i wydajna metoda wykrywania obiektów, która traktuje wykrywanie obiektów jako problem regresji. YOLO dzieli obraz na siatkę komórek i przewiduje ograniczenia pudełek oraz prawdopodobieństwa klas dla każdej komórki. YOLO jest znacznie szybsze niż inne metody, takie jak R-CNN czy Faster R-CNN, zachowując jednocześnie wysoką dokładność detekcji.

4.6 SSD (Single Shot MultiBox Detector)

SSD to kolejna szybka i wydajna metoda wykrywania obiektów, która łączy predykcje ograniczeń pudełek i prawdopodobieństw klas w jednym etapie. SSD osiąga porównywalną dokładność detekcji do Faster R-CNN, ale jest znacznie szybsze.

5 Dwuetapowe i jednoetapowe detektory obiektów

W artykule "Object Detection in 20 Years: A Survey", autorzy przedstawili różnice między dwuetapowymi a jednoetapowymi detektorami obiektów. Oto najważniejsze aspekty:

5.1 Dwuetapowe detektory

Dwuetapowe detektory, takie jak R-CNN, Fast R-CNN czy Faster R-CNN, składają się z dwóch głównych etapów:

- **Generowanie propozycji regionów:** W pierwszym etapie, detektor generuje propozycje regionów, które mogą zawierać obiekty. W przypadku Faster R-CNN, używany jest Region Proposal Network (RPN) do generowania propozycji regionów.
- **Klasyfikacja i regresja:** W drugim etapie, detektor klasyfikuje obiekty w wygenerowanych regionach i dokonuje regresji współrzędnych pudełek ograniczających.

Dwuetapowe detektory zwykle osiągają lepszą dokładność niż jednoetapowe detektory, ale są wolniejsze.

5.2 Jednoetapowe detektory

Jednoetapowe detektory, takie jak YOLO czy SSD, łączą generowanie propozycji regionów i klasyfikację obiektów w jednym etapie. W przypadku YOLO, obraz jest dzielony na siatkę komórek, a detektor przewiduje ograniczenia pudełek oraz prawdopodobieństwa klas dla każdej komórki. W przypadku SSD, detektor łączy predykcje ograniczeń pudełek i prawdopodobieństw klas w jednym etapie.

Jednoetapowe detektory są zwykle szybsze niż dwuetapowe detektory, ale mogą osiągać nieco niższą dokładność.

5.3 Podsumowanie

Główna różnica między dwuetapowymi a jednoetapowymi detektorami obiektów polega na podejściu do generowania propozycji regionów i klasyfikacji obiektów. Dwuetapowe detektory, takie jak R-CNN, Fast R-CNN czy Faster R-CNN, dzielą te zadania na dwa etapy, co pozwala osiągnąć lepszą dokładność, ale kosztem wydajności. Jednoetapowe detektory, takie jak YOLO czy SSD, łączą te zadania w jednym etapie, co pozwala na szybsze działanie, ale może prowadzić do nieco niższej dokładności.

6 Podstawowe elementy systemu wykrywania obiektów: szczegółowy przegląd

W artykule "Object Detection in 20 Years: A Survey" autorzy przedstawili podstawowe elementy systemu wykrywania obiektów. W niniejszym artykule naukowym podzielimy te elementy na kilka sekcji, aby przedstawić je bardziej szczegółowo.

6.1 Cechy obrazu

Wcześniejsze metody wykrywania obiektów opierały się na ręcznie projektowanych cechach, takich jak cechy Haar'a, histogramy zorientowanych gradientów (HOG), SIFT czy SURF. W przypadku głębokich sieci neuronowych, cechy te są uczone automatycznie przez sieć.

6.1.1 Cechy Haar'a

Cechy Haar'a są prostymi cechami obrazu, takimi jak różnice jasności między prostokątnymi obszarami obrazu. Są one wykorzystywane w detektorze Viola-Jones, który był jednym z pierwszych skutecznych algorytmów wykrywania obiektów, szczególnie twarzy.

6.1.2 Histogramy zorientowanych gradientów (HOG)

HOG to metoda wykrywania obiektów oparta na analizie gradientów obrazu. HOG dzieli obraz na małe, przesuwne okna, a następnie oblicza histogramy gradientów w każdym oknie. Histogramy te są następnie używane jako cechy do trenowania klasyfikatora, takiego jak maszyna wektorów nośnych (SVM).

6.1.3 SIFT (Scale-Invariant Feature Transform)

SIFT to algorytm wykrywania i opisywania cech obrazu, który jest niezmienniczy na skalę, rotację i zmiany oświetlenia. SIFT identyfikuje punkty charakterystyczne na obrazie, a następnie oblicza deskryptory oparte na gradientach obrazu wokół tych punktów. Deskryptory te mogą być używane do dopasowywania punktów charakterystycznych między różnymi obrazami, co jest przydatne w wykrywaniu obiektów.

6.1.4 SURF (Speeded-Up Robust Features)

SURF to usprawniona wersja SIFT, która jest szybsza i bardziej odporna na zmiany oświetlenia. Podobnie jak SIFT, SURF identyfikuje punkty charakterystyczne na obrazie i oblicza deskryptory oparte na gradientach obrazu wokół tych punktów.

6.2 Klasyfikatory

Klasyfikatory są używane do rozróżniania obiektów należących do różnych klas. Wcześniejsze metody wykorzystywały klasyfikatory takie jak maszyny wektorów nośnych (SVM). W przypadku głębokich sieci neuronowych, klasyfikacja jest zintegrowana z procesem uczenia cech.

6.3 Propozycje regionów

W przypadku dwuetapowych detektorów, takich jak R-CNN, Fast R-CNN czy Faster R-CNN, generowanie propozycji regionów jest kluczowym elementem systemu wykrywania obiektów. W przypadku jednoetapowych detektorów, takich jak YOLO czy SSD, propozycje regionów są generowane jednocześnie z klasyfikacją obiektów.

6.4 Optymalizacja i uczenie

Systemy wykrywania obiektów muszą być optymalizowane pod względem dokładności i szybkości działania. W przypadku głębokich sieci neuronowych, optymalizacja polega na uczeniu wag sieci, co pozwala na automatyczne uczenie się cech i klasyfikacji obiektów.

6.5 Zbiory danych i metryki

W celu oceny i porównania różnych metod wykrywania obiektów, potrzebne są odpowiednie zbiory danych i metryki. W artykule przedstawiono różne zbiory danych używane w badaniach nad wykrywaniem obiektów oraz metryki, takie jak mAP (średnia precyzja) czy IoU (Intersection over Union).

6.6 Techniki przyspieszające

W celu zwiększenia szybkości działania systemów wykrywania obiektów, stosuje się różne techniki przyspieszające, takie jak RoI Pooling, Region Proposal Network (RPN) czy techniki oparte na sprzęcie, takie jak GPU.

6.7 Wieloskalowe wykrywanie

Jednym z podstawowych wyzwań technicznych jest wykrywanie obiektów o różnych rozmiarach i proporcjach. Istnieją różne techniki radzenia sobie z tym problemem:

- Piramidy cech i ślizgające okna
- Wykrywanie z wykorzystaniem propozycji obiektów
- Wykrywanie oparte na wielu zakresach i rozdzielczościach
- Wykrywanie niezależne od ramki

6.8 Wzmacnianie kontekstowe

Kontekst wizualny może poprawić wykrywanie. Istnieją trzy rodzaje wzmacniania kontekstowego:

- Wykorzystanie kontekstu lokalnego
- Wykorzystanie kontekstu globalnego
- Interaktywne konteksty

6.9 Inne ważne elementy

Istnieją także inne istotne elementy, takie jak:

- Ekstrakcja trudnych negatywów (hard negative mining)
- Funkcje straty i regresja ramki ograniczającej
- Tłumienie nie-maksymalne (NMS)
- Przyspieszanie wykrywania za pomocą obrazów integralnych, przekształceń w dziedzinie częstotliwości, itp.

7 Nowoczesne techniki i ulepszenia w wykrywaniu obiektów

7.1 Ulepszenia w architekturach sieci neuronowych

Nowoczesne techniki wykrywania obiektów opierają się na zaawansowanych architekturach sieci neuronowych, takich jak ResNet, Inception czy DenseNet. Te architektury pozwalają na efektywniejsze uczenie się cech obrazów i klasyfikacji obiektów, co prowadzi do lepszej dokładności detekcji.

7.2 Techniki augmentacji danych

Augmentacja danych to proces tworzenia nowych danych treningowych poprzez modyfikację istniejących danych, np. przez obracanie, przeskalowanie czy zmianę oświetlenia obrazów. Techniki augmentacji danych pozwalają na zwiększenie ilości danych treningowych i poprawę generalizacji modeli wykrywania obiektów.

7.3 Transfer learning

Transfer learning to technika, która pozwala na wykorzystanie wiedzy nabytej przez modele na jednym zadaniu do rozwiązania innych, podobnych zadań. Transfer learning pozwala na szybsze uczenie się modeli i osiągnięcie lepszych wyników na nowych zadaniach.

7.4 Techniki uczenia z małymi ilościami danych

W przypadku wykrywania obiektów, często występuje problem z dostępnością dużych ilości danych anotowanych. Nowoczesne techniki, takie jak uczenie z małymi ilościami danych (few-shot learning) czy uczenie z aktywnym wykorzystaniem danych (active learning), pozwalają na efektywniejsze uczenie się modeli wykrywania obiektów, nawet gdy dostępne są tylko niewielkie ilości danych.

7.5 Techniki detekcji obiektów w czasie rzeczywistym

Szybkość działania detektorów obiektów jest kluczowa w wielu zastosowaniach, takich jak analiza obrazów w czasie rzeczywistym czy systemy autonomicznych pojazdów. Nowoczesne techniki, takie jak YOLO czy SSD, pozwalają na wykrywanie obiektów w czasie rzeczywistym, zachowując jednocześnie wysoką dokładność detekcji.

7.6 Integracja z innymi technologiami

Nowoczesne techniki wykrywania obiektów są coraz częściej integrowane z innymi technologiami, takimi jak analiza obrazów, sztuczna inteligencja czy teledetekcja. Integracja tych technologii pozwala na uzyskanie bardziej kompleksowych wyników i lepszego zrozumienia analizowanych danych.

8 Nowoczesne techniki przyspieszające i optymalizacja

Istnieje kilka technik stosowanych w celu przyspieszenia sieci neuronowych wykrywających obiekty.

8.1 RoI Pooling

RoI (Region of Interest) Pooling to technika stosowana w detektorach dwuetapowych, takich jak Fast R-CNN, która pozwala na szybsze przetwarzanie obrazów i wykrywanie obiektów. RoI Pooling redukuje wymiary regionów zainteresowania do stałego rozmiaru, co pozwala na efektywniejsze przetwarzanie przez sieć neuronową.

8.2 Region Proposal Network (RPN)

Region Proposal Network (RPN) to technika stosowana w detektorze Faster R-CNN, która pozwala na generowanie propozycji regionów zawierających obiekty. RPN jest zintegrowany z głęboką siecią neuronową, co pozwala na szybsze generowanie propozycji regionów i wykrywanie obiektów.

8.3 Techniki oparte na sprzęcie

Wykorzystanie sprzętu, takiego jak GPU (Graphics Processing Unit), pozwala na przyspieszenie obliczeń związanych z wykrywaniem obiektów. Dzięki zastosowaniu GPU, głębokie sieci neuronowe mogą być trenowane i wykonywać detekcję obiektów znacznie szybciej niż przy użyciu tradycyjnych procesorów CPU.

8.4 Techniki optymalizacji

W celu poprawy wydajności i dokładności detektorów obiektów, stosuje się różne techniki optymalizacji, takie jak regularyzacja, optymalizacja hiperparametrów czy uczenie z małymi ilościami danych. Optymalizacja pozwala na lepsze dopasowanie modeli do danych treningowych i testowych, co prowadzi do lepszej generalizacji i wydajności detektorów.

8.5 Techniki przyspieszające dla jednoetapowych detektorów

Jednoetapowe detektory, takie jak YOLO czy SSD, łączą generowanie propozycji regionów i klasyfikację obiektów w jednym etapie, co pozwala na szybsze działanie. W przypadku YOLO, obraz jest dzielony na siatkę komórek, a detektor przewiduje ograniczenia pudełek oraz prawdopodobieństwa klas dla każdej komórki. W przypadku SSD, detektor łączy predykcje ograniczeń pudełek i prawdopodobieństw klas w jednym etapie.

9 Przyszłe kierunki badań

W artykule "Object Detection in 20 Years: A Survey", autorzy przedstawili kilka przyszłych kierunków badań w dziedzinie wykrywania obiektów. Oto niektóre z nich:

9.1 Ulepszanie istniejących modeli

Jednym z kierunków przyszłych badań jest dalsze ulepszanie istniejących modeli wykrywania obiektów, zarówno pod względem dokładności, jak i szybkości działania. Możliwe ulepszenia obejmują optymalizację architektury sieci neuronowych, zastosowanie nowych technik uczenia maszynowego czy wykorzystanie nowych funkcji aktywacji.

9.2 Rozwój nowych technik adnotacji obrazu

Opracowywanie nowych technik adnotacji obrazu, które pozwolą na lepsze uczenie się modeli wykrywania obiektów, jest kolejnym kierunkiem przyszłych badań. Przykłady takich technik to zastosowanie pudełek ograniczających 3D, wykrywanie linii czy wykorzystanie semantycznych map głębi.

9.3 Integracja z innymi dziedzinami

Integracja wykrywania obiektów z innymi dziedzinami, takimi jak analiza obrazów, sztuczna inteligencja czy teledetekcja, może prowadzić do nowych zastosowań i lepszego zrozumienia analizowanych danych. Współpraca z innymi dziedzinami może również prowadzić do opracowania nowych technik wykrywania obiektów, które będą lepiej dostosowane do specyficznych zastosowań.

9.4 Wykorzystanie nowych technologii

Wykorzystanie nowych technologii, takich jak uczenie federacyjne, uczenie nienadzorowane czy uczenie przez wzmacnianie, może prowadzić do opracowania nowych metod wykrywania obiektów, które będą lepiej dostosowane do różnych zastosowań i scenariuszy. Wykorzystanie tych technologii może również pozwolić na lepsze wykorzystanie dostępnych danych i zasobów obliczeniowych.

9.5 Współpraca między detektorami

W przyszłości możemy spodziewać się większej współpracy między różnymi detektorami obiektów, co pozwoli na lepsze wykorzystanie ich mocnych stron i osiągnięcie lepszych wyników. Możliwe jest również opracowanie nowych metod łączenia wyników z różnych detektorów, co pozwoli na lepsze wykrywanie obiektów w różnych warunkach i scenariuszach.

10 Bibliografia

References

- [1] Zhengxia Zou, Keyan Chen, Zhenwei Shi, Member, IEEE, Yuhong Guo, and Jieping Ye, Fellow, IEEE, "Object Detection in 20 Years: A Survey"
- [2] Girshick, R. (2015). Fast R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 1440-1448).
- [3] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In *Advances in Neural Information Processing Systems* (pp. 91-99).
- [4] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 779-788).
- [5] Viola, P., & Jones, M. (2001). Rapid Object Detection using a Boosted Cascade of Simple Features. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 1-9).
- [6] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 770-778).

- [7] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going Deeper with Convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1-9).
- [8] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely Connected Convolutional Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 4700-4708).
- [9] Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How Transferable are Features in Deep Neural Networks? In *Advances in Neural Information Processing Systems* (pp. 3320-3328).
- [10] Vinyals, O., Blundell, C., Lillicrap, T., Wierstra, D., & others. (2016). Matching Networks for One Shot Learning. In *Advances in Neural Information Processing Systems* (pp. 3630-3638).
- [11] Settles, B. (2009). Active Learning Literature Survey. *University of Wisconsin-Madison Department of Computer Sciences*.
- [12] Girshick, R. (2015). Fast R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 1440-1448).
- [13] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems* (pp. 1097-1105).
- [14] McMahan, H. B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2016). Communication-Efficient Learning of Deep Networks from Decentralized Data. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics* (pp. 1273-1282).
- [15] Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., ... & Petersen, S. (2015). Human-Level Control through Deep Reinforcement Learning. *Nature*, 518(7540), 529-533.