

Wizja komputerowa - Segmentacja semantyczna i instancyjna - wykład 15

Adam Szmigielski

aszmigie@pjwstk.edu.pl

materials: *ftp(public) : //aszmigie/WMA*

Segmentacja obrazu

Segmentacja obrazu może przybierać różne formy. Niekoniecznie musi to mieć coś wspólnego z granicami obiektów.

Ogólnie rzecz biorąc, istnieją trzy grupy zadań segmentacji obrazu:

- Segmentacja semantyczna,
- Segmentacja instancji,
- Segmentacja panoptyczna.

Segmentacja semantyczna

- Segmentacja semantyczna to technika, która umożliwia powiązanie każdego piksela obrazu cyfrowego z etykietą klasy,
- Jest to również uważane za zadanie klasyfikacji obrazu na poziomie pikseli, ponieważ obejmuje rozróżnianie obiektów na obrazie.
- Segmentacja semantyczna klasyfikuje piksele obrazu jednej lub więcej klas zamiast obiektów świata rzeczywistego, których nie można interpretować semantycznie.
- Jest to trudne zadanie w wizji komputerowej - każdy piksel jest klasyfikowany zamiast obiektów, jak to ma miejsce w wykrywaniu obiektów.

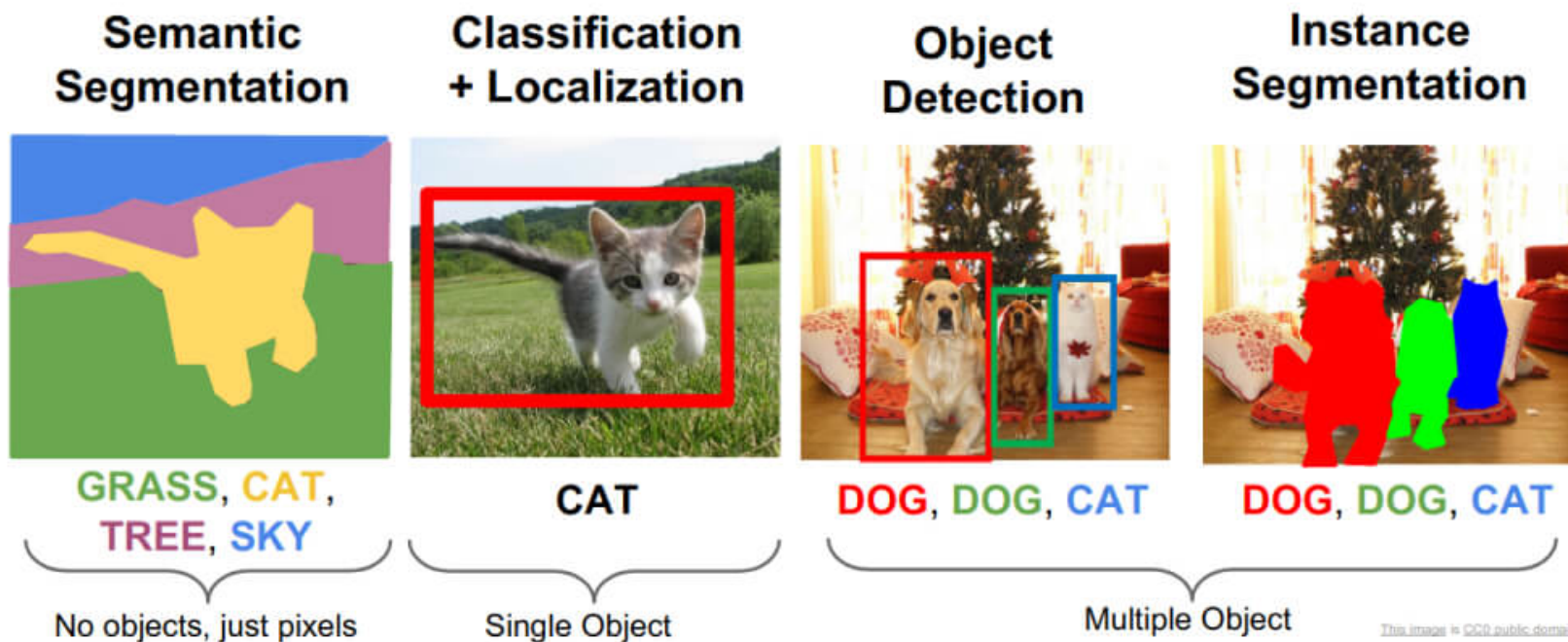
Jak działa segmentacja semantyczna?

Segmentacja semantyczna ma na celu wyodrębnienie cech przed użyciem ich do utworzenia odrębnych kategorii na obrazie.

Wymagane kroki są następujące:

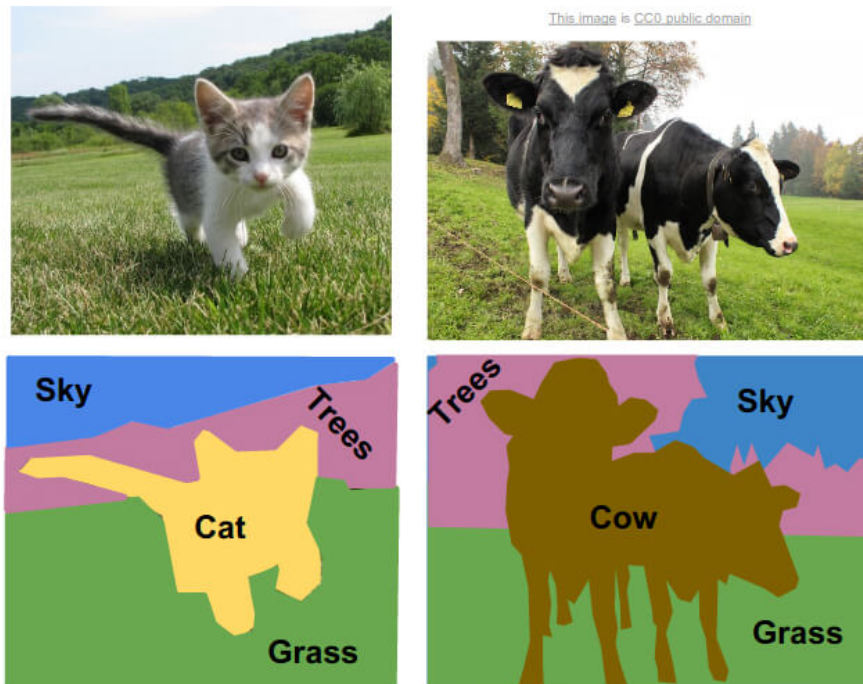
- Analizuj dane szkoleniowe w celu sklasyfikowania określonego obiektu na obrazie,
- Utwórz sieć segmentacji semantycznej, aby zlokalizować obiekty i narysuj wokół nich ramkę ograniczającą,
- Naucz sieć segmentacji semantycznej grupowania pikseli w zlokalizowanym obrazie, tworząc maskę segmentacji.

Segmentacja semantyczna



Cechą definiującą segmentację semantyczną, która odróżnia ją od segmentacji instancji, jest to, że nie rozróżnia ona różnych obiektów należących do tej samej klasy.

Segmentacja semantyczna

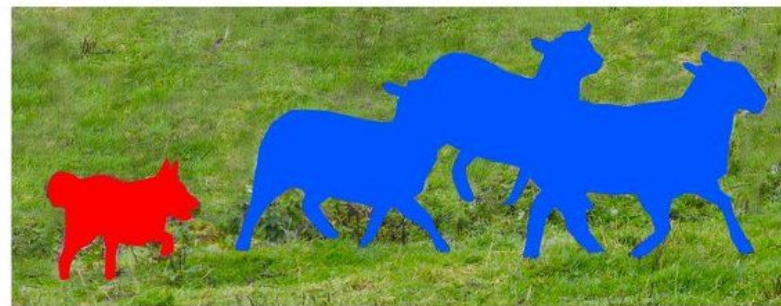


Na powyższym obrazku, zamiast dwóch różnych instancji “krowy”, otrzymujemy rodzaj kropelki pikseli należących do klasy “krowa”.

Semantyczna segmentacja poprzez rozpoznawanie



Image Recognition



Semantic Segmentation

Segmentacja instancyjna

- Segmentacja instancyjna wykrywa wszystkie instancje klasy z dodatkową funkcją wyznaczania oddzielnych instancji dowolnej klasy segmentu.
- Jest również określany jako zawierający funkcje wykrywania obiektów i segmentacji semantycznej.
- Segmentacja instancyjna ma bogatszy format wyjściowy, ponieważ tworzy mapę segmentów dla każdej kategorii i instancji tej klasy.

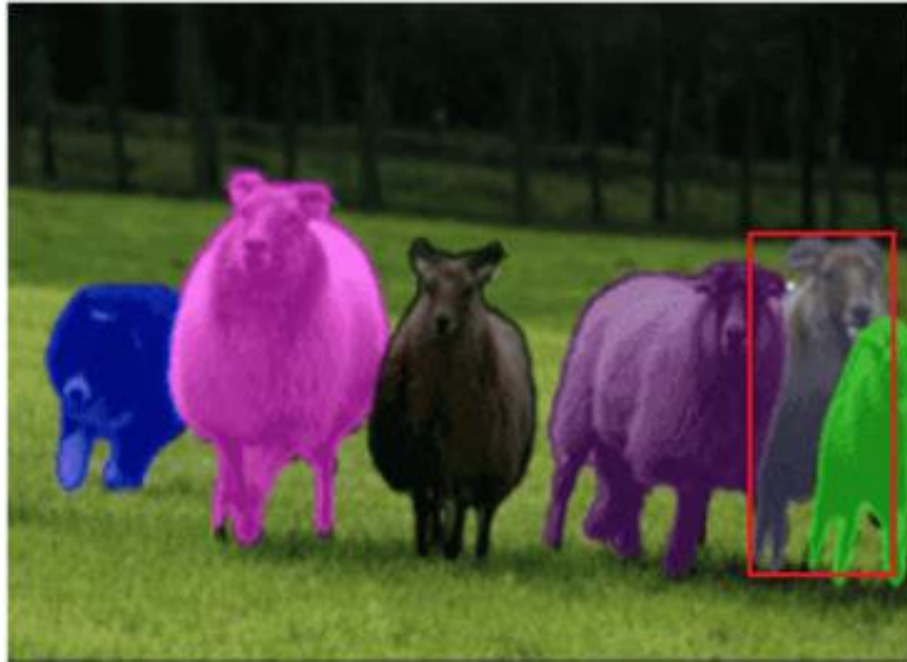
Jak działa segmentacja instancyjna?

Segmentacja instancyjna polega na identyfikowaniu granic obiektów na poziomie szczegółowych pikseli, co sprawia, że jest to złożone zadanie.

Segmentacja instancji obejmuje 2 istotne części:

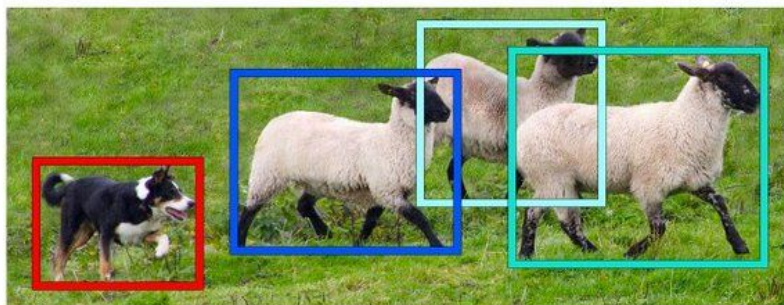
- **Wykrywanie obiektów:** Po pierwsze, uruchamia wykrywanie obiektów, aby znaleźć wszystkie obwiednie dla każdego obiektu na obrazie,
- **Segmentacja semantyczna:** Po znalezieniu wszystkich ramek ograniczających używa modelu segmentacji semantycznej wewnątrz każdego prostokąta.

Instance segmentation

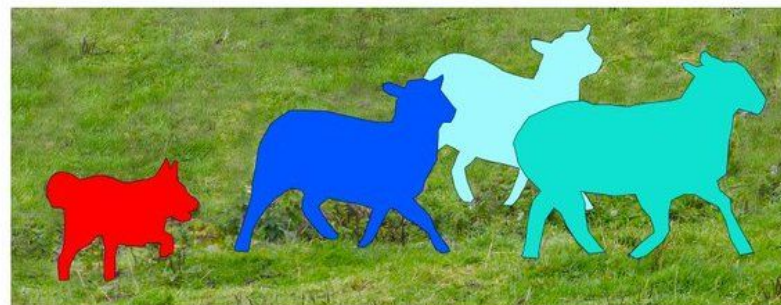


Segmentacja instancyjna jest bardzo podobna do segmentacji semantycznej. Jediną różnicą jest to, że rozróżnia różne obiekty tej samej klasy.

Segmentacja instancyjna poprzez wykrywanie



Object Detection



Instance Segmentation

Segmentacja panoptyczna



- Segmentacja semantyczna jest odpowiednia do etykietowania niezliczonych obiektów, takich jak “niebo” lub “ocean” lub po prostu obiektów, które nas interesują, klastrów, takich jak “liście” lub trawa,
- Segmentacja instancyjna dobrze nadaje się do zrozumienia obiektów policzalnych,
- Wszystkie obiekty obrazu można podzielić w ten sposób na “klasy policzalne” lub “klasy niepoliczalne”,
- Aby poradzić sobie z obydwoma typami, z pomocą przychodzi segmentacja panoptyczna. Zasadniczo łączy oba zadania w jedno.

Segmentacja semantyczna a segmentacja instancyjna

Segmentacja semantyczna

- Dla każdego piksela na danym obrazie wykrywa kategorię obiektu, do której należy, przy czym wszystkie etykiety kategorii obiektów są znane modelowi
- Segmentacja semantyczna nie pozwala na rozróżnienie różnych instancji w tej samej kategorii,
- Najpierw następuje wykrycie celu, a następnie każdy piksel jest etykietowany.
- Lista zestawów danych typu open source: Stanford Background Dataset, Microsoft COCO Dataset, MSRC Dataset, KITTI Dataset i Microsoft AirSim Dataset

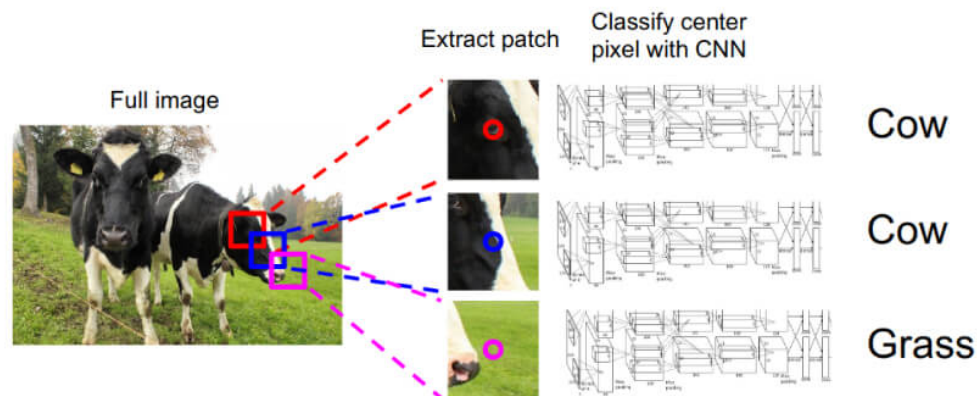
Segmentacja semantyczna a segmentacja instancyjna

Segmentacja instancyjna

- Dla każdego piksela na danym obrazie identyfikuje instancję obiektu, do której należy. Zanurza się głębiej niż segmentacja semantyczna i rozróżnia dwa obiekty z tymi samymi etykietami.
- Segmentacja instancji umożliwia rozróżnienie różnych instancji tej samej kategorii,
- Jest to hybryda adnotacji wykrywania celu i segmentacji semantycznej.
- Lista zbiorów danych typu open source: LiDAR Bonnetal Dataset, HRSID (High-Dimension SAR Images Dataset), SSDD (SAR Ship Detection Dataset), Pascal SBD Dataset i iSAID (A Large Scale Aerial Images Dataset).

Konwolucyjne sieci neuronowe do segmentacji semantycznej

Semantic Segmentation Idea: Sliding Window

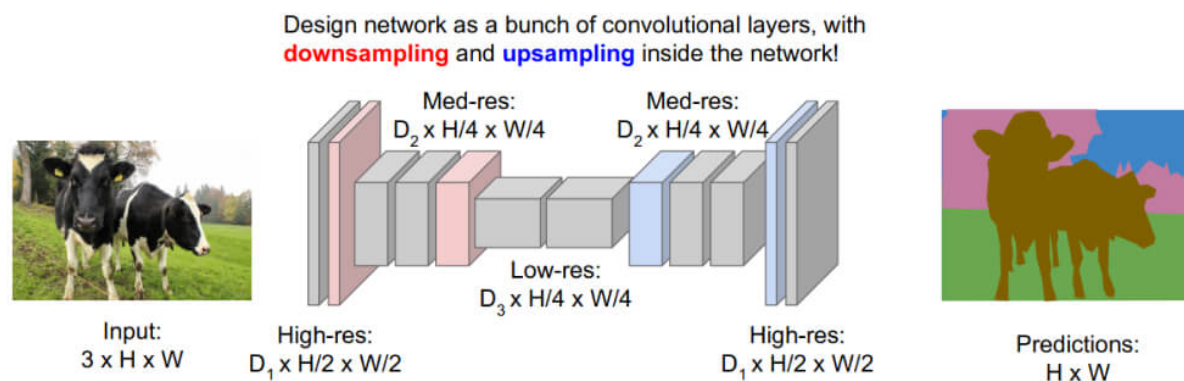


Farabet et al, "Learning Hierarchical Features for Scene Labeling," TPAMI 2013
Pinheiro and Collobert, "Recurrent Convolutional Neural Networks for Scene Labeling", ICML 2014

- Jest bardzo nieefektywny, ponieważ nie wykorzystuje ponownie funkcji, które są wspólne dla łąt.
- Nie wykorzystuje efektywnie informacji przestrzennych, a różne rozmiary przesuwanych okien mogą prowadzić do różnej dokładności w zależności od domeny, z której pochodzi obraz.

Próbkowanie w górę i w dół

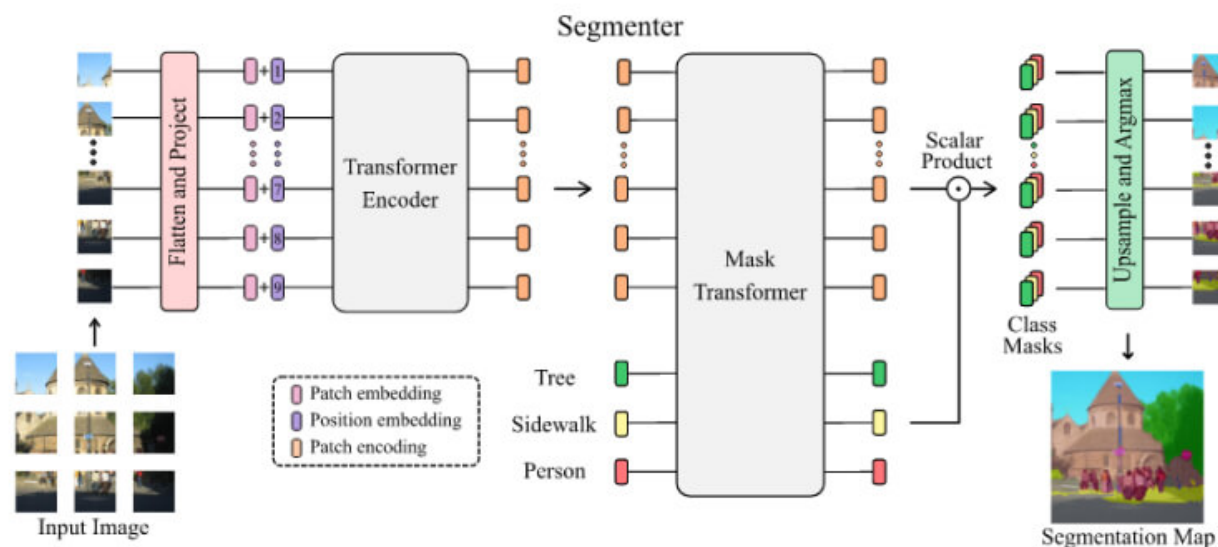
Semantic Segmentation Idea: Fully Convolutional



Long, Shelhamer, and Darrell, "Fully Convolutional Networks for Semantic Segmentation", CVPR 2015
Noh et al, "Learning Deconvolution Network for Semantic Segmentation", ICCV 2015

- W ten sposób drastycznie zmniejszymy liczbę parametrów i operacji potrzebnych do wytrenowania naszej sieci neuronowej.
- Aby wykonać te operacje próbkowania w górę i w dół, użyjemy warstw łączących.

Segmentator: Architektura



- Segmenter wykorzystuje architekturę kodera-dekodera, aby najpierw wyodrębnić funkcje, a następnie utworzyć mapę segmentacji z dokładnością do piksela.

Algorytmy segmentacji obrazu

- Do rozwiązywania tego zadania zaprojektowano sporo algorytmów, takich jak algorytm Watershed, progowanie obrazu, grupowanie K-średnich, metody partycjonowania wykresów itp.
- Zaproponowano również wiele architektur głębokiego uczenia (takich jak w pełni połączone sieci do segmentacji obrazów), ale model DeepLab firmy Google dał najlepsze wyniki do tej pory.

Google's DeepLab

Model DeepLab zasadniczo składa się z dwóch etapów:

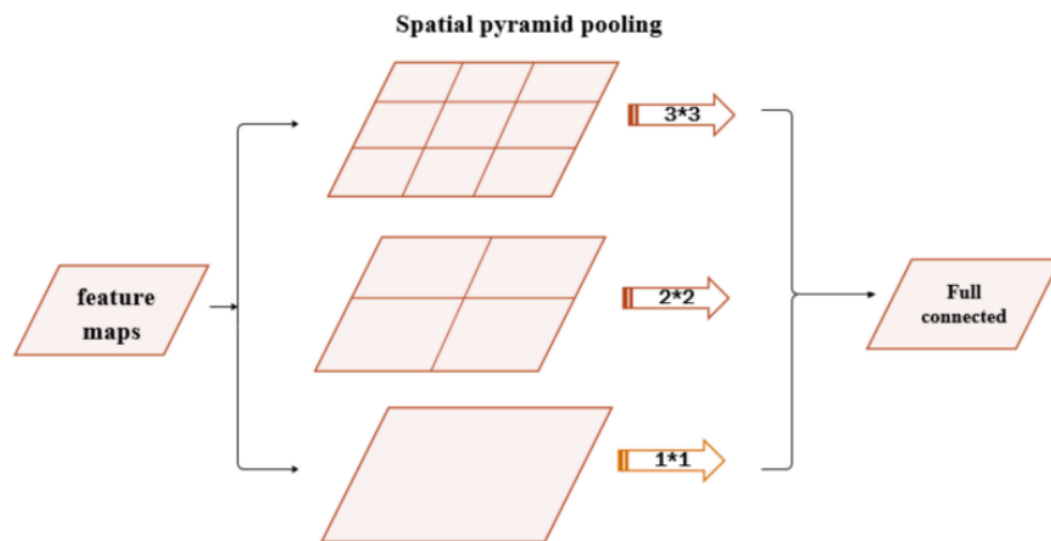
- Faza kodowania: Celem tej fazy jest wydobywanie istotnych informacji z obrazu,
- Faza dekodowania: Informacje wyodrębnione w fazie kodowania są tutaj używane do rekonstrukcji danych wyjściowych o odpowiednich wymiarach.

Fazy kodowania i dekodowania

Architektura DeepLab opiera się na połączeniu dwóch popularnych architektur sieci neuronowych:

- Łączenie piramidy przestrzennej
- Sieci koder-dekoder

Piramida przestrzenna



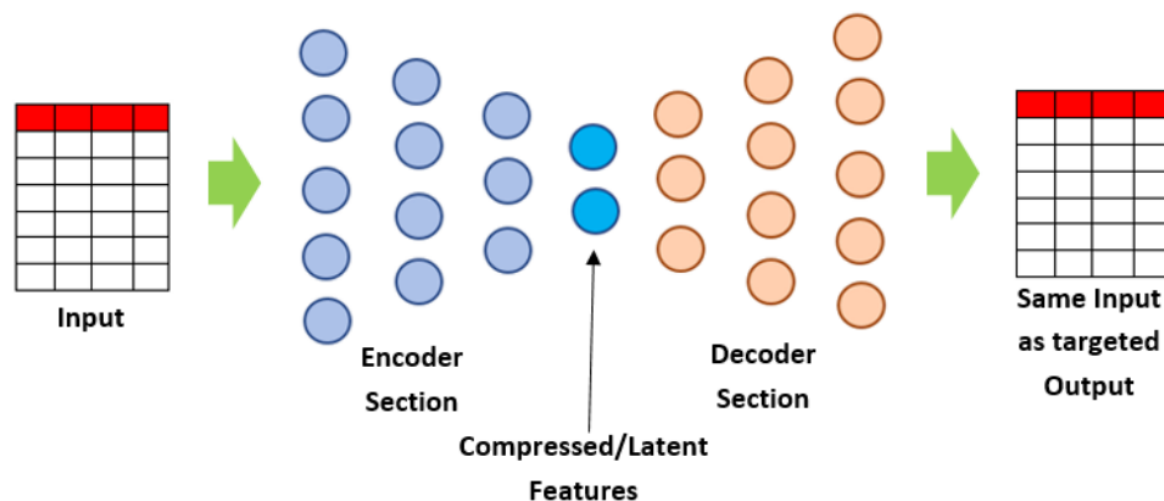
Aby upewnić się, że model jest odporny na zmiany rozmiaru obiektów podczas pracy z CNN.

- Ten problem można rozwiązać za pomocą przestrzennych sieci łączenia piramid. Wykorzystują one wiele skalowanych wersji

danych wejściowych do uczenia, a tym samym przechwytyują informacje w wielu skalach.

- Sieci łączenia piramidy przestrzennej są w stanie zakodować wieloskalowe informacje kontekstowe. Odbywa się to poprzez sondowanie nadchodzących obiektów lub operacje łączenia z różnymi szybkościami i efektywnym polem widzenia.
- Sieci łączenia piramidy przestrzennej zazwyczaj używają równoległych wersji tej samej podstawowej sieci do trenowania danych wejściowych w różnych skalach i łączenia funkcji na późniejszym etapie.

Sieci koder-dekoder



- W tym przypadku dobrze sprawdzają się sieci Encoder-Decoder. Uczą się przekształcać dane wejściowe w gęstą formę, której można użyć do przedstawienia wszystkich informacji wejściowych.