# Samenvatting paper Kenan Yildirim

<https://proceedings.neurips.cc/paper/2013/file/f7cade80b7cc92b991cf4d2806d6bd78-Paper.pdf>

Het artikel bespreekt het belang van objectdetectie voor een volledig begrip van beelden en hoe vooruitgang in machine learning-modellen en objectrepresentaties hebben geleid tot meer precieze en gedetailleerde objectherkenning. Deep Neural Networks (DNN's) worden benadrukt als een krachtig machine learning-model voor objectdetectie, dat in staat is complexe modellen te leren zonder de noodzaak van handmatige ontwerpfuncties. Het artikel beschrijft een formulering die in staat is om meerdere objecten in een afbeelding te detecteren en hun locatie te bepalen met behulp van DNN's en een eenvoudige bounding box-inferentie. De methode is eenvoudig toe te passen op een breed scala aan klassen en levert betere detectieprestaties op over een breed scala van objecten. De methode wordt gedemonstreerd aan de hand van state-of-the-art-detectieresultaten op de Pascal VOC-uitdaging.

Het deformable part-based model is een veel bestudeerd paradigma voor objectdetectie dat een set van discriminatief getrainde delen combineert in een stermodel genaamd picturale structuur. Andere op delen gebaseerde modellen, zoals het And/Or graph model, zijn gemakkelijker te interpreteren dan DNN's. Neurale netwerken zijn minder interpreteerbaar dan op delen gebaseerde modellen, maar kunnen efficiënter worden toegepast op objectdetectie door regressie op de volledige afbeelding uit te voeren. Deze aanpak wordt gebruikt in het beschreven model, in tegenstelling tot andere aanpakken die lokale of semi-lokale classificatie gebruiken.

Een methode wordt voorgesteld om precieze object localisatie te bereiken met behulp van deep neural network (DNN) gegenereerde masks. Hoewel deze aanpak in staat is om hoogwaardige masks te genereren, zijn er verschillende extra uitdagingen. Zo kan één objectmasker niet voldoende zijn om objecten die naast elkaar geplaatst zijn te onderscheiden. Bovendien genereren ze masks die veel kleiner zijn dan de grootte van de originele afbeelding. Om deze problemen aan te pakken, worden meerdere masks gegenereerd om robuuste localisatie te bereiken, die elk het hele object of een deel daarvan vertegenwoordigen. Verder wordt een score gebruikt om de overeenkomst tussen bounding boxes en masks te meten om de nauwkeurigste resultaten te verkrijgen.

Het netwerk heeft een eenvoudige structuur zonder gladheid of convolutionele lagen, maar heeft veel trainingsdata nodig. Er worden duizenden monsters van elke afbeelding genomen om de mask generator en de classifier te trainen. Voor de training van de mask generator worden negatieve monsters gedefinieerd als die geen bounding box overlappen en positieve monsters als die ten minste 80% van een bounding box overlappen. Voor de training van de classifier worden negatieve monsters gedefinieerd als die minder dan 0,2 Jaccard-similarity hebben met objecten en positieve monsters moeten ten minste 0,6 similarity hebben en worden geëtiketteerd met de klasse van het object waarmee ze het meest overeenkomen. Het hele netwerk, inclusief convolutionele lagen, wordt fijngestemd door stochastic gradient met ADAGRAD te trainen om de leersnelheid van de lagen automatisch te schatten.