Jarn Vaerewijck en Jarno De Smedt

Wij hebben gekozen voor de contrastive learning opdracht op basis van de CIFAR-10 dataset, de bedoeling is om een model te trainen om te herkennen welke afbeeldingen bij elkaar horen. Dit doen we door 2 versies van dezelfde afbeelding te geven en het model te trainen dat deze hetzelfde zijn.

Als pretext task voor deze opdracht moesten we aangepaste versies van dezelfde afbeelding maken, hiervoor gebruiken we torchvision.transforms om een willekeurige crop, een willekeurige horizontale flip en colorjitter uit te voeren. Ten laatste transformeren we de afbeelding in een tensor. Dit wordt 2 keer gedaan voor elke afbeelding om zo de 2 gerelateerde afbeeldingen te genereren.

Er wordt een custom dataset klasse gemaakt die voor elk datapunt 2 deze 2 nieuwe datapunten teruggeeft.

Als model hebben we een SimCLR model gemaakt op basis van een pretrained Resnet model (voor computer vision).

Deze bestaat uit een encoder, een projection head en een loss functie.

De encoder is een CNN die zich vooral moet focussen op het leren van de structuren in de afbeeldingen. We gebruiken een pretrained resnet als basis voor dit gedeelte. Om deze te kunnen gebruiken verwijderen we de classificatie laag van het model en voegen een projectionhead toe.

De projectionhead is een kleiner netwerk die de bevindingen van de encoder omzet naar een output die begrijpbaar is voor de loss functie. De encoder (Resnet) heeft namelijk heel veel dimensies. De projectionhead is dus alleen nodig tijdens de training en kan bij het testen achtergelaten worden.

Ten laatste hebben we de loss functie, voor SimCLR is dit het contrastive loss algoritme NT-Xent. Dit algoritme zoekt om de gelijkheden tussen de geleerde structuren van de paren zo hoog mogelijk te krijgen en de verschillen tussen de andere afbeeldingen zo hoog mogelijk te krijgen.

Wij hebben t-SNE gebruikt om te zien hoe het algoritme de afbeeldingen groepeert. Dit is gedaan op een algoritme dat voor 5 epochs getraind is. We zien dat de encoder nog veel fouten maakt maar wel al bepaalde dingen kan beginnen groeperen. We zien bijvoorbeeld een grote groep van trucks met 2 kleine groepen van auto’s. Daarnaast is een groep boten dat ook wat doordringt in de boten. Er zijn de beginsels van een groep paarden en er is al een duidelijke groep kikkers met daar tussen een groepje herten. Ook het begin van een groep honden is al zichtbaar.

Voor de andere afbeeldingen zijn er nog steeds problemen: de vliegtuigen en vogels zijn een groep op dezelfde plaats aan het vormen en de katten zijn zowat overal.

Als het algoritme langer getraind wordt zouden we waarschijnlijk betere resultaten zien. Het Resnet model is heel groot dus training duurt heel lang.

A colorful circle with text

Description automatically generated with medium confidence