**Fundamentals of Computer Vision**

**Azure AI Vision och Bildbehandling – En Översikt**

**Azure AI Vision** erbjuder kraftfulla verktyg för att analysera, bearbeta och förstå bilder genom **machine learning** och **computer vision**-tekniker. Bilder tolkas som **pixel arrays**, där varje pixel representeras av numeriska värden. Genom att använda filter och neurala nätverk kan AI extrahera och tolka information från bilder för att utföra uppgifter som **bildklassificering, objektigenkänning och bildtextgenerering**.

**Hur bilder representeras digitalt**

**Pixel Arrays och Färgkanaler**

För en dator är en bild en **matris av pixelvärden**. I en **grayscale image** har varje pixel ett värde mellan **0 (svart) och 255 (vitt)**. Exempel:

0 0 0 0 0

0 255 255 255 0

0 255 255 255 0

0 0 0 0 0

Färgade bilder använder tre **kanaler**:

* **Röd (Red)**
* **Grön (Green)**
* **Blå (Blue)**

Varje färgkomponent har sitt eget värde, och genom att kombinera dessa skapas färgtoner. Exempel:

* **Lila:** **Red: 150, Green: 0, Blue: 255**
* **Gul:** **Red: 255, Green: 255, Blue: 0**

**Bildbehandling och Filter**

**Konvolutionsfilter**

En vanlig teknik inom **image processing** är att använda **filter kernels** – små matriser som modifierar pixelvärden för att skapa effekter som **kanthöjning, oskärpa och skärpning**.

Exempel på en **3x3 laplace filter kernel**:

-1 -1 -1

-1 8 -1

-1 -1 -1

Filtret appliceras över bilden genom en **konvolution**, där varje pixel beräknas baserat på sina grannar. Exempel på **edge detection** med denna metod används i datorseende för att identifiera objekt i bilder.

**Machine Learning för Computer Vision**

**Convolutional Neural Networks (CNNs)**

CNN-modeller används för att extrahera **features** från bilder och identifiera objekt. En **CNN** använder lager av **filter** för att analysera och klassificera bilder.

**Hur CNN fungerar:**

1. Bilddata matas in i **neuralt nätverk**.
2. **Filter extraherar features** och skapar **feature maps**.
3. **Feature maps** omvandlas till en **enkel array av numeriska värden**.
4. Data bearbetas av en **fully connected neural network**.
5. Modellen förutsäger en **klassetikett** (t.ex. *äpple, banan, apelsin*).

**Träning av CNN-modeller**

* **Filtervärden (weights)** börjar slumpmässigt.
* Modellen tränas med **etiketterade bilder**.
* **Backpropagation** justerar vikterna baserat på skillnaden mellan förutsägelse och korrekt klass.
* Efter flera träningscykler (epochs) kan modellen **klassificera nya bilder**.

**Transformers och Multi-Modal AI för Computer Vision**

**Transformers i Computer Vision**

Traditionellt har **CNNs** varit dominerande inom **computer vision**, men **transformers**, som används inom **natural language processing (NLP)**, har blivit alltmer populära.

Transformers fungerar genom att:

1. **Representera data som embeddings** – numeriska vektorer som fångar relationer mellan ord eller bildfeatures.
2. **Lära sig sammanhang genom attention layers** – liknande hur **GPT-modeller** bearbetar text.

**Multi-Modal Models**

Multi-modala modeller, som **Microsoft Florence**, tränas på både **text- och bilddata** för att förstå relationen mellan bilder och text. Exempel på användningsområden:

* **Bildklassificering**
* **Objektigenkänning**
* **Bildbeskrivning (captioning)**
* **Tagging av bildinnehåll**

**Florence är en foundation model**, vilket betyder att den kan **anpassas till specifika uppgifter** utan att behöva tränas från grunden.

**Azure AI Vision – Funktioner och Användningsområden**

**Azure AI Vision Services**

Microsofts **Azure AI Vision** tillhandahåller **förtränade datorseendemodeller** baserade på **Florence**. Den används för:

1. **Optical Character Recognition (OCR)** – Extraherar text från bilder.
2. **Bildbeskrivning (Captioning)** – Genererar textbaserade bildbeskrivningar.
3. **Objektigenkänning** – Identifierar vanliga objekt i bilder.
4. **Bildtaggning** – Associerar metadata och nyckelord med bilder.

**Analysera Bilder med Azure AI Vision**

**1. Optical Character Recognition (OCR)**

Azure AI Vision kan läsa text i bilder, exempelvis en **näringsdeklaration** på en produkt:

**Inmatad bild:**

*Näringsfakta på en matförpackning*

**AI-genererad textutvinning:**

Serving size: 1 bar (40g)

Calories: 190

Total Fat: 13g

Cholesterol: 0mg

**2. Bildbeskrivning (Captioning)**

AI kan tolka och beskriva en bild:

**Inmatad bild:**  
*Man som hoppar med skateboard*

**AI-genererad beskrivning:**  
*"A man jumping on a skateboard."*

**3. Objektigenkänning**

Azure AI Vision identifierar objekt och ger en **confidence score**:

* **Skateboard (90.40%)**
* **Person (95.5%)**

Bounding box-koordinater returneras också, vilket möjliggör visuell representation av **identifierade objekt** i en bild.

**4. Bildtaggning**

AI kan generera **relevanta bildtaggar**, exempelvis för skateboardbilden:

* **Sport (99.60%)**
* **Person (99.56%)**
* **Skateboard (86.87%)**
* **Extreme sport (88.35%)**

**Träning av Anpassade Bildmodeller**

**1. Bildklassificering**

Användare kan träna en modell för att klassificera bilder. Exempel:

* *Apple, Banana, Orange*

**2. Objektigenkänning**

AI kan tränas att **lokalisera och identifiera** objekt i bilder genom **bounding boxes**.

Exempel:

* *Identifiera flera frukter i en bild och markera deras positioner.*

**Sammanfattning**

* **Bilder hanteras som pixelarrays**, där färgade bilder har **RGB-kanaler**.
* **Filter används för att förbättra eller analysera bilder**, exempelvis **edge detection**.
* **CNN-modeller används för att identifiera objekt och klassificera bilder**.
* **Transformers och multi-modal AI-modeller**, som **Florence**, kan tolka både bild- och textdata.
* **Azure AI Vision tillhandahåller förtränade datorseendemodeller** och möjliggör **OCR, bildbeskrivning, objektigenkänning och bildtaggning**.
* **Anpassade modeller kan tränas för specifika uppgifter**, såsom **bildklassificering och objektidentifiering**.

**Azure AI Vision** gör det möjligt att snabbt bygga och implementera avancerade datorseendelösningar utan att behöva utveckla modeller från grunden.