

GEMOOTHEIDS- BEPALING VAN TOMAAT

Gemoetheid is een vormeigenschap van tomaten, voornamelijk belangrijk bij de vleestomaten. Met deep learning, een vorm van artificiële intelligentie (AI), kunnen tomaten worden geclassificeerd naargelang verschillende gemoetheidsklassen. Foto's van de kroonzijde geven hiervoor betere resultaten dan deze van de evenaarszijde. Het gevonden algoritme kan even goed classificeren als de keurders.

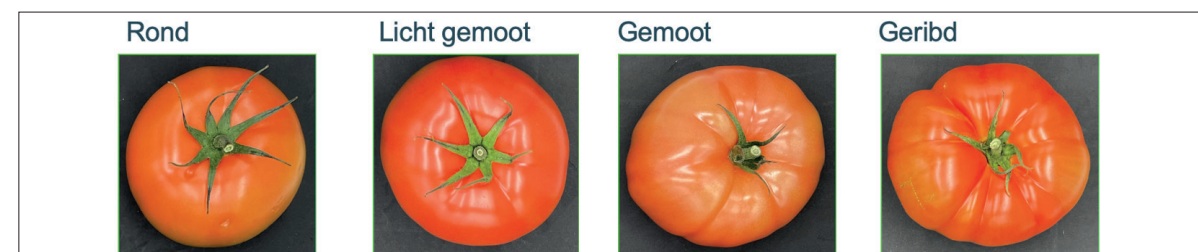
DEEP LEARNING
BIEDT
MOGELIJKHEDEN
VOOR
GEMOOTHEIDS-
BEPALING

WAT IS GEMOOTHEID?

In de supermarkt is er een groot aanbod terug te vinden aan tomaten. Een uiterlijk kenmerk zoals de vorm van de tomaten speelt dan ook een grote rol in de keuze van de consument. De vruchten dienen er goed uit te zien. Specifiek voor vleestomaten speelt hierin dus niet enkel de ideale kleur een rol. De vorm van een tomaat is niet enkel gekarakteriseerd door de grootte en de hoogte- en breedteverhoudingen van de tomaat maar ook door de zogenaamde gemoetheid. Dit is een vormeigenschap bepaald door de hoeveelheid en de diepte van inkepingen en groeven op de kroonzijde van de tomaat en hoe ver deze groeven kunnen doorlopen op de zijkanten. Gemoetheid speelt vooral een rol bij vleestomaten waar deze vorm gewenst is. Anderzijds is het belangrijk voor andere segmenten van losse tomaten dat de gemoetheid afwezig blijft. Het bepalen van de gemoetheid is daarom een onderdeel van het segmentatie-onderzoek.

EEN AI ALGORITME OM GEMOOTHEID TE BEOORDELEN

Om de gemoetheid op te meten worden er op dit moment visuele beoordelingen uitgevoerd door keurders en onderzoekers van de proeftuinen. We vroegen ons echter af of dit niet sneller kan met behulp van beeldverwerkingstechnieken deep learning, een vorm van artificiële



Figuur 1: Fotoreeks van links naar rechts: een perfect ronde tomaat, een licht gemote tomaat, een gemote tomaat en een geribde tomaat.

intelligentie (AI). In AI worden computerprogramma's gebruikt om data, foto's of afbeeldingen te analyseren. In ons geval willen we de foto's classificeren. Het computerprogramma bestaat uit een algoritme, dit is in essentie een set van instructies en regels waarmee het een complex probleem tracht op te lossen. Specifiek in dit geval bestaat de input uit foto's gelinkt aan een gemoetheidsschaal. Het algoritme 'leert' hiermee welke tomaat welke gemoetheid heeft van een bestaande set van foto's.

VEEL BEELDEN NODIG...

In een later stadium kan het algoritme dan gebruikt worden om nieuwe foto's te beoordelen. Het is belangrijk dat elke foto onder dezelfde omstandigheden wordt genomen, want lichtinval en schaduw kunnen de interpretatie beïnvloeden. Bovendien is er zeer veel ruwe data (foto's met de bijhorende beoordeling van gemoetheid) nodig om een goed AI algoritme te kunnen bouwen. We verzamelden een duizendtal foto's van tomaten, elk met een verschillende gradatie van gemoetheid.

Een foto kan meerdere tomaten bevatten die elk door het computerprogramma worden herkend. Elke tomaat wordt apart geselecteerd en verwerkt tot een foto van één enkele tomaat. Elke individuele tomaat delen we daarna in naargelang vier verschillende klassen: rond, licht gemoot, gemoot en geribd (Figuur 1). Met al deze gelabelde foto's trainen we dan een deep learning

algoritme. Daarna volgt er een validatiestap om het algoritme te controleren op nauwkeurigheid. Het beste algoritme kan dan aan de slag met onbekende data, of nieuwe foto's die het vooraf nog nooit eerder heeft geanalyseerd. Deze foto's kunnen dan worden ingevoerd, waarna het algoritme elke tomaat op de foto aan een klasse zal toewijzen.

KROONZIJD VOORSPELT GEMOOTHEID BETER DAN EVENAARZIJD

Gemoetheid is een vormeigenschap die zich over de gehele tomaat bevindt. De groeven lopen door tot op of in sommige gevallen zelfs voorbij het evenaarsvlak. Een foto kan echter maar een beeld geven in twee dimensies. Daarom testten we of we voor de bepaling van de gemoetheid enkel foto's van de kroonzijde of de evenaarszijde nodig hadden, en of dat er eventueel een verschil zou zijn tussen beoordeling van de algoritmes voor de verschillende zijden. Foto's van zowel evenaarsvlak als van het kroonvlak (Figuur 2) werden gebruikt om een gelijke werkwijze een algoritme te bouwen. De resultaten en de nauwkeurigheid van beide bekomen algoritmes werd daarna vergeleken. Hieruit bleek dat het beste algoritme voor het evenaarsvlak minder nauwkeurig is dan het beste algoritme van de kroonzijde. De kroonzijde is dus beter geschikt om tomaten te classificeren naar gemoetheid. Dit is niet zo vreemd vermits gemoetheid grotendeels te zien is op de kroonzijde en in sommige gevallen zich niet zo sterk doorzet naar het evenaarsvlak.



Figuur 2: Foto's van enkele tomaten met het kroonvlak (links) en het evenaarsvlak (rechts) die werden gebruikt om algoritmes op te bouwen

HET ALGORITME DOET HET EVEN GOED ALS DE KEURDERS

Het beste algoritme was in staat om 75,3% van de tomaten in de juiste klasse in te delen. De gemote tomaten zijn in het algemeen iets moeilijker te onderscheiden van de licht gemote tomaten. De ronde en geribde tomaten worden het gemakkelijkst geclassificeerd. Het algoritme werd ook getoetst ten opzichte van keurders. De voorspelling van het algoritme is hetzelfde als deze van een gemiddelde visuele beoordeling. We vonden wel verschillen tussen het algoritme en individuele keurders. Dat is ook logisch: we zoeken immers naar een objectieve instrumentele oplossing omdat de keurders onderling kunnen verschillen en niet even objectief zijn voor deze beoordeling. Je kan dan ook niet verwachten dat de nieuw ontwikkelde methode, als ze beter is, nog overeenkomt met de individuele resultaten van de keurders, maar wel met hun gemiddelde resultaat. Het algoritme evenaart dus de visuele keuring en is operatoronafhankelijk: eender wie het toestel bedient zal dezelfde gemoetheid voor een bepaalde partij tomaten bekomen. Bij een keurderbeoordeling is dat nu niet zo. Alhoewel er nog verder gesleuteld dient te worden aan het algoritme om het meer nauwkeurig te maken, doet het huidige algoritme het niet slechter dan de keurders. Het ontbreken van een duidelijk definieerbaar onderscheid tussen de klassen zou een mogelijke oorzaak kunnen zijn van de gevonden verschillen. Er is een betere definitie nodig voor de klassen van gemoetheid. Verder zou het uitbreiden van de dataset ook nog voor een verbetering zorgen van de resultaten.

USE OF AN AI-ALGORITHM TO DETERMINE GROOVEDNESS OF TOMATO

An important tomato shape descriptor is called 'groovedness', a specific descriptor used to describe the amount and depth of the grooves or notches mainly found on the crown side of the tomato. A 4-level scale was created for classification: perfectly round, 'lightly grooved', 'grooved' or 'ribbed'. Images were collected and used to train a deep learning algorithm. Models using images taken from the crown side performed better than models using images from the equatorial plane. A model accuracy of 75,3% was found. Comparing the AI-algorithm with the expert panel showed that the algorithm was equally good at classifying images into the different classes, with the added benefit that the AI-algorithm is user independent.

Dit onderzoek valt binnen de basiswerking van het VCBT met financiering door de Vlaamse Overheid (Departement Landbouw en Visserij), het Verbond van de Belgische Tuinbouwcoöperaties (VBT) en de provincie Vlaams-Brabant.