English Summary

We are living in an information era where the amount of image and video data increases exponentially. It is important to develop intelligent visual understanding systems to satisfy our need for searching information of interest. An important example of such a system that, with the current increasing concern for public security, is urgently required, is an automated person Re-Identification (ReID) system. A ReID system aims at understanding video streams to identifying target pedestrians across a network of disjoint surveillance cameras at distinct locations and time.

To enable ReID systems to meet the so-called open-world challenges, we explore three themes that are challenging yet practical in real application scenarios.

The first theme of this thesis is to address the lifelong learning challenges that the ReID task faces in an open world. Therefor we propose and define a new ReID task called Lifelong ReID (LReID), which aims to develop a ReID system that can continuously learn in order to address new tasks from new data while preserving knowledge learned from previous tasks. Unlike the learning process of humans, training previously trained deep networks on new data leads the networks to forget what was learned before. To address these challenges, we propose a new framework, inspired by the cognitive process of the human brain. The framework aims to solve the LReID challenges from the perspective of learnable knowledge, normalization of knowledge, and knowledge transfer. It uses three novel components: an adaptive knowledge accumulation (AKA) component, meta-reconcilement normalization (MRN) and a ranking consistency knowledge distillation (RCD). AKA aims at establishing a dynamic knowledge graph to memorize previously-learned knowledge which is adaptively updated when new knowledge arrives. This enables the model to explicitly accumulate knowledge and prevent forgetting. MRN is motivated by the synaptic plasticity in our brain, it tries to reconcile the discrepancies between old and new knowledge by learning a more generalized knowledge representation that is shared across different tasks. RCD proposes a new ReID-specific method to transfer the retrieval ranking ability from the previously-trained model to the current training model. Thus consolidating old knowledge while providing enough flexibilities to learn new knowledge. The three components are shown to be able to efficiently collaborate with each other and achieve new state-of-the-art performance in the LReID task.

The second theme of this thesis is to study the capacity of the ReID model to adapt to target domains in the case that target domain labels are not available. This is called unsupervised domain adaptation (UDA). To perform the ReID task under the UDA setting we propose a new transductive domain invariant embedding network (DIEN) with a semi-supervised triplet loss function, which enables the model to learn from both the labeled source-domain data and the unlabeled target-domain data. We also propose a recurrent top-down attention mechanism in order to further improve the generalization ability of the model. Experimental results show that our method achieves competitive performance compared with the state-of-the-art methods.

For the third theme, we consider a practical cross-modality application scenario, where the ReID systems are required to perform RGB-Infrared image retrieval for day and night re-identification. Although existing RGB-Infrared retrieval systems provide a feasible solution when identifying pedestrians under poor illumination, e.g., at night or in dark scenarios, their effectiveness is limited due to the heterogeneity of the data. To narrow this heterogeneity gap, we proposed a new variational auto-encoder with a triplet swap reconstruction strategy. Guided by a mixture of Gaussian distributions, and a standard Gaussian distribution, this strategy allows us to disentangle identity-discriminable and identity-ambiguous information, respectively. By employing the identity-discriminable information shared across RGB and Infrared image pairs, the cross-modality retrieval performance of the model is significantly improved. Furthermore, classical dissimilarity representations use a set of distances between an object and other objects to form a unique representation of the object. Inspired by this, we proposed a learnable dissimilarity metric by incorporating graph convolutional networks to bridge the heterogeneity gap between the different modalities. Finally we develop an end-to-end fusion framework to improve the accuracy of cross-modality ReID and extract precise depth information from a single image. It is shown that our proposed methods achieve the new state-of-the-art performance in the RGB-Infrared ReID tasks.

We conducted thorough experiments to verify the efficacy of the proposed methods for the three themes. The results demonstrate significant improvements over various baselines and state-of-the-art methods. Thus this thesis provides important novel contributions, insights, and findings for the research community and future applications in the field of visual understanding and person re-identification systems.

Nederlandse Samenvatting

We leven in een informatietijdperk. De hoeveelheden beeld en video data nemen exponentieel toe. Daarom is het belangrijk om intelligente systemen te ontwikkelen die deze data begrijpen en het mogelijk maakt om te kunnen zoeken naar relevante informatie. Met de huidige toenemende zorgen om de publieke veiligheid is het automatisch identificeren van personen in verschillende video streams uit een netwerk van vele verschillende surveillance camera's op verschillende tijden en plaatsen een belangrijk probleem.

Om er voor te zorgen dat de ontworpen automatische systemen voor de (her-)identificatie van personen de uitdagingen van een open wereld aan kunnen onderzoeken we drie thema's die tegelijkertijd uitdagend en van groot belang zijn in de verschillende praktische toepassingsscenario's.

Het eerste thema van dit proefschrift betreft een van de uitdagingen van het (her-)identificeren van personen (de HID-taak) in een open wereld: levens lang leren. We definiëren dit als een nieuwe HID-taak namelijk levenslange HID (LHID). Deze taak heeft als doel om een HID-systeem te ontwikkelen dat in staat is om continu te leren en zo nieuwe taken aan kan leren met behulp van nieuwe data, terwijl de kennis van eerder geleerde taken behouden blijft. In tegenstelling tot hoe mensen leren, leidt het trainen van een diep neuraal netwerk op nieuwe data er vaak toe dat eerder geleerde kennis wordt vergeten. Om deze uitdaging het hoofd te bieden introduceren we een nieuw raamwerk dat geïnspireerd is door het cognitieve proces van het menselijk brein. Het raamwerk probeert de LHID-uitdaging op te lossen vanuit het perspectief van leerbare kennis, het normaliseren van kennis, en kennisoverdracht. Het gebruikt hiervoor drie nieuwe componenten: een component voor de adaptieve verzameling van kennis (Adaptive Knowledge Accumulation (AKA)), een component voor normalisatie bij het reconciliëren van meta data (Meta-Reconcilement Normalization (MRN)), en een ranking consistente kennis-distillatie (RCD). AKA heeft als doel een dynamisch kennisnetwerk op te stellen welke eerder geleerde kennis representeert en welke adaptief wordt geüpdatet indien nieuwe kennis zich aandient. Dit zorgt ervoor dat het model expliciet kennis kan vergaren waarbij het vergeten van kennis wordt voorkomen. De inspiratie voor MRN is de synaptische plasticiteit van ons brein. Het tracht discrepanties tussen eerder vergaarde kennis en nieuwe

kennis met elkaar te verenigen door het aanleren van een meer gegeneraliseerde kennisrepresentatie die gebruikt kan worden voor een grotere verscheidenheid aan taken. RCD heeft als doel om de retrieval ranking van het eerder getrainde model over te dragen naar het huidige model. Hierbij wordt oude kennis geconsolideerd terwijl er genoeg flexibiliteit is om nieuwe kennis aan te leren. We tonen aan dat de drie componenten efficiënt samen kunnen werken en zo state-of-the-art resultaten voor de LHID-taak behalen.

Het tweede thema van dit proefschrift is de studie naar het vermogen van het HID-model om zich aan te passen aan de doel-domeinen wanneer labels van het doel-domein niet beschikbaar zijn. Dit wordt domein adaptatie zonder toezicht (Unsupervised Domain Adaptation (UDA)) genoemd. Om de HID-taak uit te voeren in een UDA-context introduceren we een nieuw transductief domein-invariant inbeddingsnetwerk met een semi-begeleide triplet verliesfunctie. Dit zorgt ervoor dat het model kan leren van zowel de gelabelde data uit het bron-domein als van de niet gelabelde data uit het doel-domein. Om het generaliserend vermogen van het model te verbeteren introduceren we tevens een recurrent top-down aandacht-mechanisme. Experimenten laten zien dat onze methode in staat is om, vergeleken met state-of-the-art methoden, competitieve resultaten te behalen.

Voor het derde thema beschouwen we een praktisch toepassingsscenario voor cross-modaliteiten. Hierbij dienen de HID-systemen in staat te zijn om RGB-Infrarood beelden te kunnen gebruiken voor dag- en nacht-HID. Alhoewel bestaande HID-systemen voor RGB-Infrarood beelden al werkbare oplossingen geven voor het identificeren van voetgangers bij slechte verlichting, bijvoorbeeld 's nachts of in donkere situaties, is de effectiviteit hiervan begrensd vanwege de heterogeniteit van de data. Om de verscheidenheid aan data het hoofd te bieden introduceren we een variabele auto-encoder met een reconstructie methode die gebruikt maakt van een drievoudige uitwisselingsstrategie. Geleid door een mix van Gaussische verdelingen, en een standaard Gaussische verdeling staat deze strategie ons toe om identiteit onderscheidende en ambigue informatie m.b.t. de identiteit van elkaar te onderscheiden. Door de identiteit onderscheidende data van RGB en Infrarood beeldparen te gebruiken zien we dat de performance van het cross-modaliteitsmodel significant verbetert. Traditioneel wordt de verzameling van afstanden van een object tot andere objecten gebruikt om tot een unieke ongelijkheidsrepresentatie van dat object te komen. Hierdoor geïnspireerd introduceren we een maat van ongelijkheid waarbij we een graaf convolutioneel netwerk gebruiken om de heterogeniteit tussen de verschillende modaliteiten aan te pakken. Tot slot ontwikkelen we een end-to-end fusie raamwerk om de precisie van cross-modaliteit-HID te verbeteren en precieze diepte-informatie uit een enkel beeld te extraheren. We laten zien dat onze voorgestelde methoden de nieuwe state-of-the-art resultaten behalen voor de RGB-infrarood HID-taak.

We voerden grondige experimenten uit om de effectiviteit van de voorgestelde methoden in ieder van de drie thema's aan te tonen. De resultaten laten significante ver-

beteringen zien ten opzichte van verscheiden baseline- en state-of-the-art methoden. Zo geeft dit proefschrift belangrijke nieuwe contributies, inzichten en bevindingen voor de onderzoeksgemeenschap en voor toekomstige toepassingen op het gebied van visual understanding en HID-systemen.