

Objektumok és hibáik felismerése

PhD tézis füzet

Amr Mohamed Abdelhameed Nagy Abdo

Témavezető: Dr. Czúni László



Villamosmérnöki és Információs Rendszerek Tanszék

Informatikai Tudományok Doktori Iskola

Pannon Egyetem

Veszprém, Magyarország

2022

1 Bevezetés

Az elmúlt években a számítógépes látás egyik legalapvetőbb és nehéz témája, az objektumfelismerés nagy figyelmet kapott. Az egyik legfontosabb feladatnak számít, amely jelentős hatással van a társadalomra, mivel számos alkalmazás épül a tárgyfelismerési technikákra, például a közlekedési táblák felismerése, a vásárlási alkalmazások, az emberfelismerés a megfigyelő rendszerekben, a közúti tárgyfelismerés az autonóm vezetésben, valamint a mindennapi tárgyfelismerés a kézi eszközökben és a robotikában. Azonban a 3D objektumfelismerés még mindig a számítógépes látás egyik központi problémája. A 3D objektumfelismerés valós környezetben, kézi eszközök vagy robotok segítségével nehéz feladat a változó nézőpontok, az eltérő 3D-ből 2D-be való vetületek, a lehetséges különböző zajok (pl. elmosódás, színtorzítás), valamint a korlátozott számítási erőforrások és memória miatt. A videófelvevételek gyakran kulcsfontosságúak számos számítógépes látásra épülő alkalmazásban, például a pilóta nélküli légi járművek (UAV) alkalmazásaiban, a feldolgozóiparban vagy a videófelügyeletben. Az objektumfelismerés célja olyan számítási modellek és technikák kifejlesztése, amelyek a számítógépes látás alkalmazásai számára az egyik legalapvetőbb információt szolgáltatják: Milyen tárgyakat látunk és pontosan hol vannak? A számítógépes látás egyik alapvető problémájaként az objektumfelismerés és detektálás számos más számítógépes látási feladat alapját képezi, mint például a vizuális ellenőrzés [2, 4], az objektumkövetés [6], a szegmentálás [1], a képek feliratozása [5] stb. Az objektumok gépi felismerése után a következő fontos lépés az anomáliák detekciója. A vizuális hibadetekciós technológia lehetővé teszi a tárgyak rendellenességeinek megkülönböztetését az emberi vizuális vizsgálatot utánzó módon. Bár elvileg minimális emberi tevékenységgel történő felügyeletet tesz lehetővé, ugyanazon megoldás alkalmazása sokféle feladatra vagy hibatípusra kihívást jelent. Az ilyen alkalmazásoknak széles

köre létezik, beleértve a közlekedési táblák hibáit, az acélfelületek hibáit, a napelemeket, az automatizált termékgyártást, a vasúti ipart, az öntést vagy a hegesztést és az egészségügyet. Az ipari elvárásoknak való megfelelés érdekében nagy szükség van a nagy teljesítmény elérésére az automatizált vizuális ellenőrzésben. Ez a munka olyan kutatási módszerekkel foglalkozik, amelyek több irányból is képesek felismerni a tárgyakat hagyományos és mély tanulási megközelítésekkel és megbízható megoldásokat nyújtanak a különböző típusú hibák hatékony felismerésére.

2 Motiváció

Az ipar 4.0 megköveteli a megbízható, pontos és gyors objektumfelismerési és vizuális ellenőrzési technológiákat meglétét a gyártási folyamatokban, amelyek gyakran az optikai információk folyamatos megfigyelésére támaszkodnak. Ezenkívül a "könnyű" objektumfelismerési módszerek nagyon hasznosak lehetnek számos alkalmazáshoz, például drónokhoz vagy hordozható okoseszközökhöz. Ezen túlmenően, mivel az automatizálás széles körben elterjedt a gyártási folyamatokban, egyre nő az igény az anomáliák automatikus észlelésére. Vagyis ha gépi intelligenciánkat betanítottuk egy adott feladatra, akkor mindig nem-nulla a valószínűsége annak, hogy olyan váratlan események történhetnek, amelyek megoldására a rendszer nincs felkészítve. Ennek a problémának egy része a "few-shot learning", ahol az új típusú hibák a lehető leghamarabb besorolásra kell hogy kerüljenek, jellemzően nagyon kevés tanítási mintával. "Incremental learning"-et kell végrehajtanunk, hiszen a korábban betanított tudást nem szabad elfelejteni. Ráadásul a teljes architektúra újratanítása erőforrásigényes lenne: a nagy mennyiségű idő, memória és feldolgozási teljesítmény jellemzően nem áll rendelkezésre a helyszínen vagy időben.

3 Tézisek összefoglalása

Ez a rész PhD kutatásom főbb, eredeti eredményeit mutatja be.

I. tézis: Objektumfelismerési technikák mély neurális hálózatok és HMM-ek használatával

Az objektumfelismerés területén csak az elmúlt néhány évben történt jelentős javulás a neurális hálózatok fejlődésével. Míg a konvolúciós neurális hálózatok nagyon hatékonyak az objektumfelismerésben, sok gyakorlati esetben még mindig fejlesztésre szorulnak. Például, ha az egyes képek feldolgozása után a teljesítmény nem kielégítő, a természetes bizonytalanságok vannak jelen (például zaj, takarások és geometriai torzulások), valamint a számítási és memóriaigények magasak. Ezenkívül a környezet tárgyaival való interakcióhoz nemcsak a konkrét vagy általános tárgyfelismerés elkerülhetetlen, hanem a pózuk meghatározása is szükséges.

1. **Egy keretrendszer javasoltam annak bemutatására, hogy a HMM-ek felhasználhatók a CNN-ek és az IMU-k adatainak kombinálására objektumfelismeréshez és pózbecsléshez.** Megmutattam, hogy a relatív pózváltozások elegendő információt adhatnak a HMM-eknek ahhoz, hogy megbecsüljék a legvalószínűbb állapotsorokat, és így megtalálják a képeken látható legvalószínűbb objektumot, és javítsák a teljesítményt. A javasolt módszer hatékonyságának bemutatására az eredményeket VGG16 hálózattal hasonlítottuk össze.
2. **A HMM alapú információk továbbfejleszthetők az aktív látás megközelítéssel, a javasolt keretrendszer neve AV-HMM-CEDD.** Megmutattam, hogy az aktív észlelés és az egymás utáni több felvételből származó információfúzió hogyan segítheti a 3Ds objektumok felismerését még akkor is, ha az objektumok részlegesen el vannak takarva. A kísérleti eredmények azt mutatják, hogy a javasolt módszerünk sokkal jobban képes kezelni

a nem betanított, takarásban lévő eseteket, ha összehasonlítjuk a DNN-ekkel. Ezenkívül a javasolt AV-HMM-CEDD keretrendszer számításilag egyszerű, minimális memóriát igényel, és más osztályozókat is használhat, nem csak a bemutatott CEDD-t.

Ezt a tézist részletesen a 2. fejezetben fejtem ki, a kapcsolódó publikációk pedig a következők: **AM1, AM2, AM3**.

II. tézis: Közlekedési táblák hibáinak észlelése

Kétségtelen, hogy a közlekedési táblák nagyon fontos részei a járművek közúti infrastruktúrájának. A közlekedési táblák nagy pontosságú észlelése és felismerése azonban még mindig megoldatlan probléma, különösen valós körülmények között. A kedvezőtlen időjárási, világítási és képképzési viszonyok mellett a közlekedési táblák nem kívánt hibái nagymértékben befolyásolhatják az ilyen rendszerek pontosságát. Elkerülhetetlen olyan rendszerek kialakítása, amelyek a közlekedési táblák állapotát figyelik, a különböző hibák felderítésével, karbantartásukat segítve.

1. **Egy új szími neurális hálózati architektúrát javasoltam a közlekedési táblák nagyszámú osztályain előforduló hibák felismerésére.** A korábban javasolt hálózatokat több réteggel kibővítettük, és az eredeti jellemzőket a különbség számítása mellett a teljesen összekapcsolt rétegeknél megtartottuk. A kísérleti eredmények azt mutatják, hogy ez a megközelítés alkalmazható a képek anomáliáinak felismerésére, jobb teljesítménnyel, mint a jól ismert tanulási megközelítések. Ezen túlmenően lehetőség van a tanításban nem résztvevő objektumok hibáinak felismerésére is.
2. **A teljesítmény javítása érdekében egy új mechanizmust javasoltam a szími hálózatok konfidencia értékeinek az SVM-mel való kombinálására, ún. "support set" képek segítségével.** Megközelítésünk előnye más, a hálózatok kombinációját

használó megoldásokkal szemben az, hogy csak egy hálózatot kell betanítani és karbantartani.

Ezt a tézist részletesen a 3. fejezet ismerteti, a kapcsolódó publikációk pedig a következők: **AM4, AM5.**

III. tézis: Klasszifikáció, Zero és Fast Few-Shot Learning alkalmazása acél felületek hibáin

Az acél a mennyisége és az alkalmazások sokfélesége szempontjából a legfontosabb fém a modern világban. Agyártási folyamat és a környezeti feltételek miatt az acélfelületeknek számos hibája lehet. A legtöbb alkalmazásban a felismerés és az osztályozás kérdései mellett a következő kulcsfontosságú problémákra kell választ adni: valós idejűség probléma, kicsi méretű hibák kérdésköre, kis mintamennyiség probléma, kiegyensúlyozatlan mintahalmaz probléma. Ráadásul az "incremental learning" során az új adatok (azaz a korábban már megismert vagy még nem látott osztályok új felvételei) időbeli fázisokban érkeznek, és az osztályozási modellünket ki kell terjesztenünk az új osztályok vagy az osztályok új mintáinak figyelembevételére.

1. **Egy új architektúrát javasoltam az EfficientNet és a randomizált hálózatok kombinálására (EffNet+RC) osztályozásra és néhány felvételes tanulásra (few-shot learning), valamint acélfelületi hibák folyamatos tanulására.** Ez az architektúra legalább egy nagyságrenddel rövidebb tanítási időt igényel, mint más osztályozók. Ezenkívül jól ismeri fel a hibákat, még ha csak néhány felvételtől is ismerjük azokat.
2. **A "zero shot learning" probléma kezelésére a szími hálózat mély architektúráját javasoltam.** Ez a hálózat alkalmas osztályozásra avagy a nem látott osztályok klaszterezésére tanítási minták nélkül.

Ezt a tézist részletesen a 4. fejezet ismerteti, a kapcsolódó publikációk pedig a következők: **AM6**, **AM7**.

4 A tudományos eredmények alkalmazásai

A dolgozatban javasolt módszerek hasznosak lehetnek az Ipar 4.0 technológiákhoz. Az Ipar 4.0 az automatizálás és az adatcsere növekvő tendenciáját testesíti meg a feldolgozóipar technológiáiban és folyamataiban, ideértve a dolgok internetét (IoT), az intelligens gyártást, a robotikát, a gépi tanulást és a mesterséges intelligenciát.

Az I. tézis eredményeit a drónok területén vagy a hordozható okoseszközöknél is alkalmazni lehetne az objektumfelismerésben, mivel kis erőforrásigényű megoldásokról van szó (kis adatmennyiség, alacsony memóriaigény). A II. tézisben javasolt módszerek segíthetnek a közlekedési táblák állapotának nyomon követésében, ami az önvezető autók jobb megbízhatóságában jelenhet meg. A McAfee kutatói például a közelmúltban könnyen becsaptak egy Teslát, hogy gyorsabban hajtson végre, miközben az autó intelligens sebességtartó funkciója bekapcsolt állapotban volt [3]. Megmutatták, hogyan tudták rávenni a Teslát, hogy egy tábla egyszerű és apró módosításával félreolvassa a sebességkorlátozást, és 35 helyett 85 mérföld/órás sebességre gyorsítson. Létrehoztunk egy adathalmazt a közlekedési táblák hibáiról, és online elérhetővé tettük laboratórium honlapján. A III. tézis eredményei közvetlenül alkalmazhatók az acéllemezek és -szalagok gyártás közbeni minőségértékelésére, mivel a hibák osztályozásában megközelítőleg 100%-os pontosságot értünk el.

Megoldásaink kifejlesztéséhez a legújabb és elterjedt szoftvertechnológiákat használtuk, például Python-t, TensorFlow-t, Keras-t és Java-t, így mások is könnyen alkalmazhatják ezeket a megoldásokat valós alkalmazásokban. Ez a munka részben a TKP2020-NKA-10, OTKA K 135729 és OTKA K 120369 projektek keretében valósult meg.

5 Publikációk

Amr Mohamed Nagy Abdo publikációi a következők:

5.1 A dolgozathoz kapcsolódó publikációk

- AM1.** Czúni László and **Amr M. Nagy**: Improving object recognition of CNNs with multiple queries and HMMs. In Twelfth International Conference on Machine Vision (ICMV), pages 1143310, 2020.
- AM2.** Czúni László and **Amr M. Nagy**: Hidden Markov models for pose estimation. In 15th International Conference on Computer Vision Theory and Applications (VISAPP), pages 1143310, 2020.
- AM3.** **Amr M. Nagy**, Metwally Rashad and Czúni László: Active multiview recognition with hidden Markov temporal support. In Signal, Image and Video Processing, pages 315–322, 2021 (**2020-IF: 2.157**).
- AM4.** **Amr M. Nagy** and Czúni László: Detecting object defects with fusioning convolutional siamese neural networks. In 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISAPP), pages 157–163, 2021.
- AM5.** **Amr M. Nagy** and Czúni László: Deep neural network models for the recognition of traffic signs defects. In 11th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), pages 725-729, 2021.
- AM6.** **Amr M. Nagy** and Czúni László: Classification and fast few-shot learning of steel surface defects with randomized network. Applied Sciences; 12(8):3967, 2022, (**2020-IF: 2.67**).
- AM7.** **Amr M. Nagy** and Czúni László: Zero-shot learning and classification of steel surface defects. In 14th International Conference on Machine Vision (ICMV), pages 386 - 394, 2021.

5.2 A dolgozathoz nem kapcsolódó publikációk

- A1.** Alejandro R. Rodriguez, **Amr M. Nagy**, Zsolt Vörösházi, György Bereczky, László Czúni: Segmentation and error detection of PV modules, In 27th International Conference on Emerging Technologies and Factory Automation, 2022.
- A2.** **Amr M. Nagy**, Ali Ahmed and Hala Helmy Zayed: Hybrid iterated Kalman particle filter for object tracking problems. In 8th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISAPP), pages 375–381, 2013.
- A3.** **Amr M. Nagy**, Ali Ahmed and Hala Helmy Zayed: Particle filter based on joint color texture histogram for object tracking. In International Image Processing, Applications and Systems Conference, pages 1–6, 2014.

5.3 További előadások

- P1.** Czúni László, **Amr M. Nagy**, M. Rashad, “About the temporal support of active object recognition”, Pannonian Conference on Advances in Information Technology (PCIT 2020)”, Veszprem, Hungary.
- P2.** Czúni László, **Amr M. Nagy**, M. Rashad, “Low complexity 3D object recognition for mobile devices based on Markovian framework (HMM)”, in: 12th Conference of Hungarian Association for Image Processing and Pattern Recognition (KÉPAF-2019), Debrecen, Hungary.
- P3.** Czúni László, **Amr M. Nagy**, M. Rashad, “Temporal models for 3D object recognition”, in: World Congress on Artificial Intelligence and Machine Learning, (WCAIML-2019), Spain.

Irodalomjegyzék

- [1] Jifeng Dai, Kaiming He, and Jian Sun. Instance-aware semantic segmentation via multi-task network cascades. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3150–3158, 2016.
- [2] Nazrul Ismail and Owais A Malik. Real-time visual inspection system for grading fruits using computer vision and deep learning techniques. *Information Processing in Agriculture*, 9(1):24–37, 2022.
- [3] Trivedi S. Povolny, S. Model hacking ADAS to save safer roads for autonomous vehicles, February 2020. [Online; posted 4-May-2022].
- [4] Andrew DH Thomas, Michael G Rodd, John D Holt, and CJ Neill. Real-time industrial visual inspection: A review. *Real-Time Imaging*, 1(2):139–158, 1995.
- [5] Qi Wu, Chunhua Shen, Peng Wang, Anthony Dick, and Anton Van Den Hengel. Image captioning and visual question answering based on attributes and external knowledge. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(6):1367–1381, 2017.
- [6] Zhengxia Zou, Zhenwei Shi, Yuhong Guo, and Jieping Ye. Object detection in 20 years: A survey. *ArXiv Preprint ArXiv:1905.05055*, 2019.