

Doktori (PhD) értekezés tézise

# Developing Robust AI Techniques in Computer Vision

Beregi-Kovács Marcell

Témavezető: Dr. Hajdu András



DEBRECENI EGYETEM

Matematika- és Számítástudományok Doktori Iskola  
Természettudományi és Műszaki Tudományi Doktori Tanács

Debrecen, 2025

# Tartalomjegyzék

<b>1. Bevezetés</b>	1
<b>2. A CNN Együttesek Diverzitásának Növelése Korrelációs Büntetéssel</b>	6
<b>3. Fizikai Alapú Kódmodellezés Inhomogén Közegben</b>	10
<b>Hivatkozások</b>	14
<b>Publikációs lista</b>	20

# 1. Bevezetés

Az elmúlt évtizedben a mesterséges intelligencia (AI) nélkülözhetetlen erővé vált a tudományos és ipari területek széles körében, különösen a számítógépes látás területén. A modern számítógépes látórendszerek kiemelkedő sikereket értek el az orvosi képalkotásban, az önvezető járművekben, a környezeti érzékelésben és a távérzékelésben. Ugyanakkor két tartós kihívással szembesülnek a valós alkalmazások során: (i) érzékenyek a környezeti változásokra és szándékos zavarásokra, valamint (ii) erősen függenek azoktól a tanító adathalmazoktól, amelyek nem mindig reprezentálják kellőképpen a ritka vagy degradált körülményeket.

Ez a disszertáció e két korlátozást két, egymástól független, de egymást kiegészítő kutatási irányon keresztül kezeli. Az első a mélytanulások osztályozó modellek robusztusságának és általánosítóképességének javítására irányul, korrelációbüntetéses együttes tanulás alkalmazásával. A második egy fizikai alapokon nyugvó algoritmust vezet be, amely a Sugárzási Transzfer Egyenlet (RTE) [1] diszkrétizált formáján alapul, és képes kültéri jelenetekben térben inhomogén köd szimulációjára. Mindkét hozzájárulás célja a látórendszerek megbízhatóságának növelése bizonytalan körülmények között, és módszertani szempontból jelentős előrelépést képviselnek saját területükön.

A konvolúciós neurális hálózatok (CNN-ek) felemelkedése példátlan pontosságot eredményezett a vizuális felismerési feladatokban. Azonban az egyes CNN-ek hajlamosak a tanító eloszlásokhoz való túllilleszkedésre, és gyengén teljesítenek eloszlásváltozás esetén. Az együttes tanulás életképes stratégiává vált e törékenységre enyhítésére, több modell predikcióinak kom-

binálásával. Az együttes ereje azonban nemcsak tagjai egyéni pontosságától, hanem azok diverzitásától is függ. Ha az együttes tagjai nem korrelált hibákat vétének, az eredményül kapott modell robusztusabb, jobban kalibrált, és megbízhatóbb lesz biztonságkritikus helyzetekben.

A gyakorlatban azonban kihívást jelent a diverzitás indukálása mélytanulós együttesekben. A hagyományos technikák – mint a különböző inicializálások, heterogén architektúrák használata, vagy adataugmentáció – gyakran korlátozott funkcionális diverzitást eredményeznek. Emellett az offline aggregáció (pl. átlagolás vagy többségi szavazás) nem ösztönzi közvetlenül a predikciók dekorrelálását a tanítás során.

Ennek kezelésére egy új, online együttes tanulási keretrendszert javasolunk, amely egy Pearson-korreláción alapuló büntetést integrál közvetlenül a veszteségfüggvénybe. Ez a büntető tag csökkenti annak valószínűségét, hogy az együttes tagjai ugyanazokat a helytelen predikciókat adják, miközben engedi az egyetértést a helyes besorolások esetén. E regularizáció beágyazásával egy közös tanítási folyamatba, a javasolt architektúra egyszerre optimalizál a pontosságra és a diverzitásra. Az így kapott együttesek javuló eloszlásfüggetlen robusztusságot, kisebb tévesztési korrelációt és csökkentett közös hibaarányokat mutatnak.

Négy benchmark adathalmazon végzett empirikus kiértékelés – természetes képek (CIFAR-10 [2], ISINI [3]) és klinikai képek (Diabetikus Retinopátia [4, 5], ISIC bőrelváltozások [6]) – megerősíti, hogy módszerünk következetesen felülmúlja a klasszikus együttes fúziós stratégiákat (pl. többségi szavazás, súlyozott átlagolás). Továbbá kimutattuk, hogy mérsékelt büntetés (pl.  $\lambda \in [0.5, 1.0]$ ) biztosítja a legjobb egyensúlyt a

besorolási pontosság és a predikciók dekorrelációja között. A módszer számításilag hatékony, architektúrafüggetlen és valós idejű rendszerekben is alkalmazható, így mind akadémiai, mind ipari környezetben bevethető.

Bár az együttes diverzitás a modelloldali robusztusságot növeli, a második kutatási irány az adatoldali korlátokat célozza. A számítógépes látás gyakorlati alkalmazásaiban a tanító adatok gyakran nem reprezentálják megfelelően az összetett vizuális torzulásokat, mint például a kód által okozott fényszórás. A hagyományos szintetikus kódgeneráló módszerek – beleértve a Koschmieder-törvényt [7] alapulókat – homogén és izotróp szórású közeggel számolnak. Ezek az egyszerűsítések megkönnyítik a számítást, de nem képesek valósághű, mélységfüggő vizuális hatások előállítására – ezáltal korlátozzák az ilyen szintetikus adatok hasznosságát a robusztus érzékelési modellek tanításában.

Válaszul kidolgoztunk egy fizikailag megalapozott szimulációs technikát, amely a RTE diszkretizált változatán alapul, és képes térben inhomogén, anizotróp módon szóró kód modellezésére. A RTE leírja, hogyan halad a sugárzás részt vevő közegen keresztül, figyelembe véve az elnyelést és a minden irányú szórást. Módszerünk diszkretizálja mind a térbeli, mind a szögbeli tartományokat, és egy rekurzív formulációt alkalmaz, amely képpontonként becsüli meg a sugárzás gyengülését és a szórásból származó hozzájárulást. A keretrendszer támogatja a mélységtudatos gyengítést, az előre irányuló szórást (Henyey-Greenstein fázisfüggvények révén), és a térben változó kihalási együtthatókat.

A számítási hatékonyság érdekében az algoritmust tensoralapú műveletekkel valósítottuk meg, amelyek lehetővé teszik

a párhuzamos végrehajtást és a gradiens alapú optimalizációt. Ez lehetővé teszi a módszer integrálását kódérzékeny számítógépes látási feladatok tanítási folyamataiba, miközben megőrzi fizikai értelmezhetőségét. A módszert tovább javítja egy kalibrációs eljárás, amely annotált ködös-felhős képpárokat használ, lehetővé téve a modell számára, hogy megtanulja a valós ködeloszlásokat.

A Koschmieder-modellen és GAN-alapú [8] ködgenerátorokkal szembeni kísérleti összehasonlítások azt mutatják, hogy megközelítésünk nagyobb realizmust és szerkezeti megőrzést ér el, amit a Fréchet Inception Distance (FID) [9], Learned Perceptual Image Patch Similarity (LPIPS) [10] és Pearson-korreláció alapján értékeltünk. Eredményeink azt sugallják, hogy a fizikavezérelt szimuláció megbízható alapot nyújthat adathalmaz-bővítéshez, végső soron javítva a látórendszerek robusztusságát csökkent látási viszonyok mellett.

Noha az együttes tanulás és a ködszimuláció elsőre különállónak tűnhet, közös céljuk van: a látórendszerek robusztusságának növelése környezeti bizonytalanság mellett. Előbbi a modell viselkedésének változékonyságát kezeli diverzitástudatos optimalizációval; utóbbi a tanító adatokat egészíti ki fizikailag megalapozott kedvezőtlen körülményekkel.

Együttesen ezek a hozzájárulások egy egységes választ alkotnak az alkalmazott AI egyik kulcskihívására: a megbízható működés biztosítása valós környezetekben. Lehetővé teszik, hogy a modellek jobban általánosítsanak különböző adathalmazokon, és alkalmazkodni tudjanak a ritka és nehéz vizuális szituációkhoz, amelyeket nehéz manuális annotációval vagy hagyományos augmentációval megragadni.

A kutatás a módszertani szigor és a gyakorlati alkalmaz-

hatóság kombinálásával továbbfejleszti a tudomány jelenlegi állását mind az algoritmikus együttesek tervezésében, mind az adatok szintézisében. Olyan eszközökkel és meglátásokkal járul hozzá, amelyek az alkalmazások széles körének hasznára válhatnak, az orvosi diagnosztikától az autonóm navigációig, és megalapozza a fizikailag inspirált modellezés és a tanulás-alapú rendszerek jövőbeli integrációját.

## 2. A CNN Együttesek Diverzitásának Növelése Korrelációs Büntetéssel

A CNN-ek a modern képosztályozás alappilléreivé váltak, mivel kiválóan képesek hierarchikus jellemzőket kinyerni vizuális adatokból. Sikereik ellenére azonban az egyes CNN-ek gyakran nehezen általánosítanak, különösen az eloszlástól eltérő bemenetek vagy zajos környezetek esetén. Az együttes tanulás [11] lehetőséget kínál e korlátok mérséklésére több modell kombinálásával, javítva ezzel a pontosságot és a robusztusságot. A hatékony együttes kulcsa azonban nemcsak tagjai egyéni teljesítményében, hanem diverzitásukban [12] rejlik — azaz abban, hogy mennyire különböző hibákat vétnek.

Ebben a kutatásban egy újszerű ensemble tanulási keretrendszert javasoltunk, amely kifejezetten elősegíti a CNN ensemble tagjai közötti sokféleséget egy Pearson korreláció-alapú büntetésen keresztül, amelyet a közös képzési veszteségbe integráltunk. A hagyományos ensemble-megközelítésekkel ellentétben, amelyek jellemzően az egymástól függetlenül képzett modellek offline aggregációjára támaszkodnak, a mi módszerünk az összes ensemble-tagot egyszerre, online módon képi, lehetővé téve az optimalizálási folyamat számára, hogy a pontosságot és a kimeneti dekorrelációt egyaránt figyelembe vegye.

A szokványos együttes technikák, mint például a többségi szavazás [13, 14, 15] vagy az átlagolás [16, 17, 18, 19, 20], gyakran nem aknázzák ki teljes mértékben az együttesekben rejlő lehetőségeket, mivel a hasonlóan betanított hálózatok kimenetei erősen korreláltak. A diverzitás mesterséges előidézésére irányuló klasszikus módszerek — például véletlenszerű inicializálás, adataugmentáció [14, 21], vagy eltérő architektúrák

alkalmazása [16, 22, 23] — közvetettek és gyakran nem ele-  
gendőek.

Ennek orvoslására a következő veszteségfüggvényt vezettük  
be:  $\mathcal{L} = \mathcal{L}_{\text{CE}} + \lambda \mathcal{L}_{\text{corr}}$  ahol  $\mathcal{L}_{\text{CE}}$  a kategóriák közötti kereszt-  
entropia veszteség függvény,  $\mathcal{L}_{\text{corr}}$  a korrelációs büntetési tag,  
 $\lambda$  pedig egy hiperparaméter, amely szabályozza a pontosság  
és diverzitás közötti egyensúlyt. A korrelációs büntetés a soft-  
max kimenetek közötti Pearson-korrelációs együttható alapján  
kerül kiszámításra, és célja a hibás predikciók közötti egyezé-  
sek büntetése, miközben engedi a helyes osztályozásokban való  
egyetértést.

Minden együttes tag egy CNN, amelynek végső teljesen  
összekapcsolt rétegét egy feladatspecifikus fejvel helyettesítet-  
tük. A hálózatok kimenetei egy közös, tanítható fúziós rétegen  
keresztül kerülnek aggregálásra, amely végül softmax aktivá-  
cióval zárul. A tanítás során a korrelációs büntetés gradiense  
a kereszt-entropia veszteséggel együtt visszaterjesztésre kerül.  
A büntetést az együttes modell minden párjára kiszámítjuk,  
biztosítva, hogy a modellek ne csak redundáns másolatai le-  
gyenek egymásnak, hanem valóban különböző mintázatokat  
tanuljanak az adatokból.

A kimeneti vektorok  $h_i$  és  $h_j$  közötti Pearson-korrelációs  
együttható a  $\rho(h_i, h_j) := \frac{\text{Cov}(h_i, h_j)}{\sigma_{h_i} \sigma_{h_j}}$  alakban adható meg, ahol  
 $\text{Cov}(h_i, h_j)$  a kovariancia,  $\sigma$  pedig a szórás. A büntetés célja,  
hogy  $\rho(h_i, h_j)$  alacsony legyen, ha mindkét predikció hibás.

A javasolt módszert négy adathalmazon teszteltük: ISINI  
és CIFAR-10 (természetes képek), valamint Diabetikus Reti-  
nopátia (DR) és ISIC (klinikai képek). Homogén (pl. több  
AlexNet [24] vagy VGG16 [25]) és heterogén (pl. AlexNet +  
ResNet50 [26] + MobileNetV2 [27]) együtteseket is vizsgál-

tunk. Az eredmények megerősítették a javasolt megközelítés hatékonyságát többféle scenárióban:

- A módszer minden esetben felülmúlta a klasszikus együttes stratégiákat (többségi szavazás, átlagolás, súlyozott szavazás) pontosság tekintetében.
- A diverzitásra vonatkozó metrikák (dupla/tripla hibák száma) szignifikánsan javultak közepes  $\lambda$  értékeknél (0.5 vagy 1.0).
- A módszer robusztusnak bizonyult különböző domének és architektúrák esetén, különösen jól szerepelt orvosi képosztályozási feladatokban.
- A  $\lambda$  paraméter lehetőséget ad a diverzitás – pontosság kompromisszum beállítására.

**Tézis 2.1.** Egy új CNN-ensemble keretrendszert hoztam létre, amelyben a taghálózatok közös tanítása egy Pearson-korrelációs büntetést is tartalmazó veszteségfüggvény segítségével valósul meg. Ez a megfogalmazás kifejezetten gátolja a redundáns hibás osztályozásokat, növeli a funkcionális diverzitást, és lehetővé teszi az ensemble end-to-end optimalizálását. A regulárizáció erősségét szabályozó  $\lambda$  paraméter expliciten biztosítja az osztályozási pontosság és a predikciós diverzitás közötti egyensúly beállítását, így a módszer a feladat- és doménspecifikus követelményekhez igazítható.

**Tézis 2.2.** A javasolt online tanítási megközelítés minden vizsgált doménben felülmúlta a klasszikus ensemble stratégiákat, mint például a többségi szavazást vagy az átlagolást, igazolva ezzel a módszer előnyét a robusztus és egymást kiegészítő modellviselkedés elősegítésében.

Összegzésként elmondható, hogy a javasolt együttes tanítási stratégia jelentős előrelépést jelent a robusztus és általánosítható CNN együttesek tervezésében. A Pearson-korrelációs büntetés tanítási folyamatba való beágyazása közvetlen, hangolható mechanizmust biztosít a funkcionális diverzitás kiképzésére. Természetes és klinikai adathalmazokon elért empirikus eredmények alátámasztják a módszer alkalmazhatóságát mind az akadémiai kutatásban, mind biztonságkritikus környezetekben.

A részletes kísérleti eredményeket és megállapításainkat a [\[28\]](#) publikációban tettük közzé.

### 3. Fizikai Alapú Ködmodellezés Inhomogén Közegekben

A környezeti viszonyok valósághű szimulációja kulcsfontosságú a számítógépes látórendszerek tanításához és értékeléséhez, különösen biztonságkritikus területeken, mint az autonóm járművek irányítása, légi navigáció vagy megfigyelés. A kedvezőtlen időjárási jelenségek közül a köd különösen nagy kihívást jelent, mivel összetett fényszórási és gyengítési hatásokat idéz elő, amelyek jelentősen rontják a vizuális információt. Bár a klasszikus, Koschmieder-törvényen alapuló ködaugmenáló módszerek széles körben elterjedtek [29, 30], ezek alapvetően korlátozottak egyszerűsítő feltételezéseik – például térbeli homogenitás és izotróp szórás – miatt.

Ezzel szemben az adatalapú megközelítések, mint például a CycleGAN [8] és a StarGAN [31, 32], képesek doméntranszformációval ködhatásokat generálni. Bár ezek a generatív elleneséges hálózatok vizuálisan meggyőző eredményeket adhatnak, hiányzik belőlük a fizikai értelmezhetőség és kontrollálhatóság. Gyakran stílusalapú hibákkal, geometriai következetlenségekkel és a fizikai ködviselkedéstől való eltérésekkel küzdenek, ahogy azt számos tanulmány is megállapította [33, 34, 35]. Emellett ezek a modellek inherensen sztochasztikusak, így kevésbé alkalmasak biztonságkritikus alkalmazásokban, ahol a reprodukálhatóság és az átláthatóság elengedhetetlen.

Ebben a kutatásban egy új ködszintézis keretrendszert javasoltunk, amely a sugárátvitel fizikáján alapul. Megközelítésünk a Sugárzási Transzfer Egyenlet (RTE) diszkretizált változatára épül – ez az egyenlet átfogóan írja le a fény és a résztvevő közegek kölcsönhatását. Az általunk javasolt mód-

szer lehetővé teszi térben inhomogén és anizotróp szórású köd szimulációját, ezáltal felülmúlva a hagyományos modellek korlátait.

Az RTE leírja a fény út menti gyengülését elnyelés és szórás révén, valamint a többi irányból történő beszóródott sugárzás hozzájárulását is. Ahhoz, hogy ezt a formulációt képtéren történő ködszintézisre alkalmassá tegyük, diszkrétizáltuk mind a térbeli, mind a szögtartományokat. Minden képpont sugár egy mélységi lépésekből álló szakaszra oszlik, a szög szerinti integrálást pedig egy mintavételezett irányhalmaz segítségével közelítjük. Az irányfüggő szórás modellezéséhez a Henyey–Greenstein fázisfüggvényt alkalmaztuk. Implementációnk képes egyetlen 2D képen belül modellezni a köd 3D geometriáját és a fény terjedését.

A komplex modell számítási kihívásainak leküzdésére egy tenzor-alapú implementációt fejlesztettünk ki, amely az  $n \times n$  méretű ködösített kép előállításának teljes számítási költségét  $\mathcal{O}(n^5)$ -ről  $\mathcal{O}(n)$ -re csökkenti. Ez a gyorsítás lehetővé teszi a keretrendszer gyakorlati alkalmazását szintetikus adathalmazok előállítására. Emellett a modell támogatja a differenciálható megjelenítést, így lehetőség nyílik a fizikai paraméterek gradiensalapú optimalizálására. Ez lehetővé teszi a köd tulajdonságainak kalibrálását valódi ködös–derült képpárok alapján, valamint a szintézis folyamat alkalmazkodását különböző légköri viszonyokhoz.

Számos kvantitatív és kvalitatív értékelést végeztünk módszerünkről, összehasonlítva azt analitikus (Koschmieder-alapú) modellel és egy modernebb GAN-alapú hálóval. A perceptuális hasonlósági metrikák – például Fréchet Inception Distance (FID) és Learned Perceptual Image Patch Similarity (LPIPS)

– alapján kimutattuk, hogy RTE-alapú módszerünk jelentősen valósághibbb és fizikailag következetesebb ködös képeket eredményez. Ezenkívül a módszer képes megőrizni a jelenetek geometriai struktúráját a mélységtérképek alkalmazásának köszönhetően, amelyeket monokuláris becslő hálózatokkal vagy sztereoó képalkotással állítottunk elő.

A javasolt keretrendszer hidat képez a fizikai modellezés és a mélytanulás között. Megbízható eszközt biztosít kiváló minőségű ködös képek generálására, és megalapozza a további kutatásokat a fizikailag megalapozott adataugmentáció, az adverszariális tesztelés és az érzékelőrendszerek bizonytalan-ságmodellezése terén.

**Tézis 3.1.** Egy új ködgenerálási keretrendszert javasoltam a diszkretizált Sugárzási Transzfer Egyenlet (RTE) alapján, amely képtérben működik, monokuláris mélységtérképeket felhasználva. Az anizotróp Henyey–Greenstein fázisfüggvény integrálásával a módszer pontosan képes modellezni az előreszóródási hatásokat, ezáltal fizikailag konzisztens és perceptuálisan realiztikus ködgenerálást tesz lehetővé, amely felülmúlja a hagyományos izotróp megközelítéseket.

**Tézis 3.2.** A diszkretizált RTE rekurzív numerikus formulációja homogén és izotróp feltételezések mellett visszavezethető a klasszikus Koschmieder-modellre, ezáltal általánosítva ezt a széles körben használt analitikus törvényt az inhomogén és anizotróp közegekre is.

**Tézis 3.3.** A javasolt RTE-alapú szimuláció mind perceptuális realizmus (LPIPS), mind eloszlásbeli illeszkedés (FID), mind pedig korreláció tekintetében következetesen felülmúlja az analitikus és a GAN-alapú ködgenerálási megközelítéseket. Ezen felül a rekurzív diszkretizációt szignifikánsan optimali-

záltam egy tensor-alapú implementációval, amely a módszert számításilag sokkal hatékonyabbá és gyakorlatilag jól skálázhatóvá teszi. A keretrendszer teljes mértékben differenciálható és hatékonyan megvalósítható mélytanulási környezetekben, így alkalmas gradiens-alapú paramétertanutásra és szintetikus adatkészletek előállítására is.

Összegzésként elmondható, hogy ez a kutatás egy fizikailag megalapozott, differenciálható kódgeneráló algoritmust nyújt, amely képes valóság-hű ködhatások szimulálására inhomogén közegben. A fizikai optika, numerikus módszerek és hatékony GPU-implementáció ötvözésével a módszer alkalmas látórendszerek tanítására és validálására változatos és kihívást jelentő vizuális körülmények között. A módszer nemcsak elméletileg megalapozott, hanem gyakorlatban is hatékony, így utat nyit a fizikailag inspirált adataugmentáció kutatása és a robusztus mesterséges intelligencia rendszerek fejlesztése felé.

A részletes kísérleti eredményeket és megállapításainkat a [\[36\]](#) publikációban tettük közzé.

## Hivatkozások

- [1] S. Chandrasekhar. *Radiative Transfer*. Dover Books on Physics. Dover Publications, 2013.
- [2] Alex Krizhevsky, Vinod Nair, and Geoffrey Hinton. Cifar-10 - Canadian Institute for Advanced Research. *MIT*, 2009. from <http://www.cs.toronto.edu/~kriz/cifar.html>.
- [3] Roy Prasun, Ghosh Subhankar, Bhattacharya Saumik, and Pal Umapada. Effects of degradations on deep neural network architectures. *ArXiv*, 07 2018. from <https://arxiv.org/pdf/1807.10108.pdf>.
- [4] Kaggle. Diabetic Retinopathy Detection, 2015. from <https://www.kaggle.com/c/diabetic-retinopathy-detection>.
- [5] Etienne Decencière, Xiwei Zhang, Guy Cazuguel, Bruno Lay, Béatrice Cochener, Caroline Trone, Philippe Gain, Richard Ordonez, Pascale Massin, Ali Erginay, Béatrice Charton, and Jean-Claude Klein. Feedback on a publicly distributed database: the messidor database. *Image Analysis & Stereology*, 33(3):231–234, Aug 2014.
- [6] Noel C. F. Codella, David Gutman, M. Emre Celebi, Brian Helba, Michael A. Marchetti, Stephen W. Dusza, Aadi Kalloo, Konstantinos Liopyris, Nabin Mishra, Harald Kittler, and Allan Halpern. Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic). In

- 2018 IEEE 15th International Symposium on Biomedical Imaging*, pages 168–172, 2018.
- [7] Hans Israël and Fritz Kasten. *KOSCHMIEDERs Theorie der horizontalen Sichtweite*, pages 7–10. VS Verlag für Sozialwissenschaften, Wiesbaden, 1959.
  - [8] Zaher Ghais. Simulating weather conditions on digital images. Master’s thesis, University of Debrecen, 2020.
  - [9] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS’17*, page 6629–6640, Red Hook, NY, USA, 2017. Curran Associates Inc.
  - [10] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.
  - [11] L.K. Hansen and Peter Salamon. Neural network ensembles. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 12:993 – 1001, 11 1990.
  - [12] Michael Perrone and Leon Cooper. When networks disagree: Ensemble methods for hybrid neural networks. *Neural networks for speech and image processing*, 08 1993.
  - [13] Yushi Chen, Ying Wang, Yanfeng Gu, Xin He, Pedram Ghamisi, and Xiuping Jia. Deep learning ensemble for

- hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(6):1882–1897, 2019.
- [14] Rodrigo Minetto, Maurício Pamplona Segundo, and Sudeep Sarkar. Hydra: An ensemble of convolutional neural networks for geospatial land classification. *IEEE Transactions on Geoscience and Remote Sensing*, 57(9):6530–6541, 2019.
- [15] Shuxian Dong, Wei Feng, Yinghui Quan, Gabriel Dauphin, Lianru Gao, and Mengdao Xing. Deep ensemble cnn method based on sample expansion for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–15, 2022.
- [16] Ishtiaque Ahmed Khan, Asaduzzaman Sajeeb, and Shaikh Anowarul Fattah. An automatic ocular disease detection scheme from enhanced fundus images based on ensembling deep cnn networks. In *2020 11th International Conference on Electrical and Computer Engineering*, pages 491–494, 2020.
- [17] Parham M. Kebria, Abbas Khosravi, Syed Moshfeq Salaken, and Saeid Nahavandi. Deep imitation learning for autonomous vehicles based on convolutional neural networks. *IEEE/CAA Journal of Automatica Sinica*, 7(1):82–95, 2020.
- [18] Wenmei Li, Haiyan Liu, Yu Wang, Zhuangzhuang Li, Yan Jia, and Guan Gui. Deep learning-based classification methods for remote sensing images in urban built-up areas. *IEEE Access*, 7:36274–36284, 2019.

- [19] Xiangrong Zhang, Wenkang Ma, Chen Li, Jie Wu, Xu Tang, and Licheng Jiao. Fully convolutional network-based ensemble method for road extraction from aerial images. *IEEE Geoscience and Remote Sensing Letters*, 17(10):1777–1781, 2020.
- [20] Baihua Zhang, Shouliang Qi, Patrice Monkam, Chen Li, Fan Yang, Yu-Dong Yao, and Wei Qian. Ensemble learners of multiple deep cnns for pulmonary nodules classification using ct images. *IEEE Access*, 7:110358–110371, 2019.
- [21] Andrey Kuehlkamp, Allan Pinto, Anderson Rocha, Kevin W. Bowyer, and Adam Czajka. Ensemble of multi-view learning classifiers for cross-domain iris presentation attack detection. *IEEE Transactions on Information Forensics and Security*, 14(6):1419–1431, 2019.
- [22] Najd Alosaimi and Haikel Alhichri. Fusion of cnn ensemble for remote sensing scene classification. In *2020 3rd International Conference on Computer Applications Information Security (ICCAIS)*, pages 1–6, 2020.
- [23] Ali Yazdizadeh, Zachary Patterson, and Bilal Farooq. Ensemble convolutional neural networks for mode inference in smartphone travel survey. *IEEE Transactions on Intelligent Transportation Systems*, 21(6):2232–2239, 2020.
- [24] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C.J. Burges, L. Bottou, and K.Q. Weinberger, editors, *Advances in Neural Informa-*

- tion Processing Systems*, volume 25. Curran Associates, Inc., 2012.
- [25] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition, 2015.
  - [26] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.
  - [27] Andrew Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv*, 04 2017.
  - [28] Balazs Harangi, Agnes Baran, Marcell Beregi-Kovacs, and Andras Hajdu. Composing diverse ensembles of convolutional neural networks by penalization. *Mathematics*, 11(23), 2023. (Q2, IF: 2.3, SJR: 0.498).
  - [29] Jean-Philippe Tarel and Nicolas Hautière. Fast visibility restoration from a single color or gray level image. In *2009 IEEE 12th International Conference on Computer Vision*, pages 2201–2208, 2009.
  - [30] Jean-Philippe Tarel, Nicolas Hautière, Aurélien Cord, Dominique Gruyer, and Houssam Halmaoui. Improved visibility of road scene images under heterogeneous fog. In *2010 IEEE Intelligent Vehicles Symposium*, pages 478–485, 2010.

- [31] Blessing Agyei Kyem, Joshua Kofi Asamoah, Ying Huang, and Armstrong Aboah. Weather-adaptive synthetic data generation for enhanced power line inspection using stargan. *IEEE Access*, 12:193882–193901, 2024.
- [32] Yating Lin, Yidong Li, Haidong Cui, and Zheng Feng. Weagan:generative adversarial network for weather translation of image among multi-domain. In *2019 6th International Conference on Behavioral, Economic and Socio-Cultural Computing (BESC)*, pages 1–5, 2019.
- [33] Fayçal Abbas and Med Babahenini. Forest fog rendering using generative adversarial networks. *The Visual Computer*, 39:1–10, 01 2022.
- [34] Rui Gong, Dengxin Dai, Yuhua Chen, Wen Li, Danda Paudel, and Luc Gool. Analogical image translation for fog generation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35:1433–1441, 05 2021.
- [35] Hanqing Zhang, Qitao Dan, and Lingfeng Wang. From point to surface: Realistic and perceptually-plausible hazy image generation with glow-diffusion. In *Pattern Recognition and Computer Vision: 7th Chinese Conference, PRCV 2024, Urumqi, China, October 18–20, 2024, Proceedings, Part X*, page 89–102, Berlin, Heidelberg, 2024. Springer-Verlag.
- [36] Marcell Beregi-Kovacs, Balazs Harangi, Andras Hajdu, and Gyorgy Gat. Generation of synthetic non-homogeneous fog by discretized radiative transfer equation. *Journal of Imaging*, 11(6), 2025. (Q1, IF: 3.3, SJR: 0.662).



Nyilvántartási szám: DEENK/422/2025.PL  
Tárgy: PhD Publikációs Lista

Jelölt: Beregi-Kovács Marcell  
Doktori Iskola: Matematika- és Számítástudományok Doktori Iskola  
MTMT azonosító: 10083987

## A PhD értekezés alapjául szolgáló közlemények

### Idegen nyelvű tudományos közlemények külföldi folyóiratban (2)

1. **Beregi-Kovács, M.**, Harangi, B., Hajdu, A., Gát, G.: Generation of Synthetic Non-Homogeneous Fog by Discretized Radiative Transfer Equation.  
*J. Imaging.* 11 (6), 1-22, 2025. EISSN: 2313-433X.  
DOI: <http://dx.doi.org/10.3390/jimaging11060196>  
IF: 3.3 (2024)
2. Harangi, B., Baran, Á., **Beregi-Kovács, M.**, Hajdu, A.: Composing Diverse Ensembles of Convolutional Neural Networks by Penalization.  
*Mathematics.* 11 (23), 1-19, 2023. EISSN: 2227-7390.  
DOI: <http://dx.doi.org/10.3390/math11234730>  
IF: 2.3

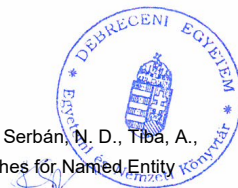
## További közlemények

### Idegen nyelvű tudományos közlemények hazai folyóiratban (1)

3. Bogacsovcics, G., Hajdu, A., Lakatos, R., **Beregi-Kovács, M.**, Tiba, A., Tomán, H.: Replacing the SIR epidemic model with a neural network and training it further to increase prediction accuracy.  
*Ann. Math. Inform.* 53, 73-91, 2021. ISSN: 1787-5021.  
DOI: <http://dx.doi.org/10.33039/ami.2021.02.003>

### Idegen nyelvű konferencia közlemények (2)

4. Bogacsovcics, G., Harangi, B., **Beregi-Kovács, M.**, Kupás, D., Lakatos, R., Serbán, N. D., Tiba, A., Tóth, J.: Assessing Conventional and Deep Learning-Based Approaches for Named Entity Recognition in Unstructured Hungarian Medical Reports.  
In: 2024 IEEE 22nd World Symposium on Applied Machine Intelligence and Informatics (SAMI). Ed.: Kovács Levente, Liberios Vokorokos, IEEE, Piscataway, 77-82, 2024. ISBN: 9798350317206





5. **Beregi-Kovács, M.**, Baran, Á., Hajdu, A.: Efficient Learning of Model Weights via Changing Features During Training.

In: 2020 IEEE 24th International Conference on Intelligent Engineering Systems (INES),

Institute of Electrical and Electronics Engineers (IEEE), Piscataway (NJ), 43-48, 2020. ISBN:

9781728110592

**A közlő folyóiratok összesített impakt faktora: 5,6**

**A közlő folyóiratok összesített impakt faktora (az értekezés alapjául szolgáló közleményekre):**

**5,6**

A DEENK a Jelölt által a Tudóstérbe feltöltött adatok bibliográfiai és tudánymetriai ellenőrzését a tudományos adatbázisok és a Journal Citation Reports Impact Factor lista alapján elvégezte.

Debrecen, 2025.06.24.



Short thesis for the degree of doctor of philosophy (PhD)

# Developing Robust AI Techniques in Computer Vision

by Marcell Beregi-Kovács

Supervisor: Dr. András Hajdu



UNIVERSITY OF DEBRECEN

Doctoral School of Mathematical and Computational Sciences  
Doctoral Council for Natural Sciences and Engineering

Debrecen, 2025

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Increasing CNN Ensemble Diversity via Correlation Penalty</b>	<b>5</b>
<b>3</b>	<b>Physically Based Fog Modeling in Inhomogeneous Media</b>	<b>9</b>
	<b>References</b>	<b>13</b>
	<b>List of publications</b>	<b>19</b>

# 1 Introduction

Artificial intelligence (AI) has become an indispensable across a broad spectrum of disciplines in the past decade, with computer vision at the heart of many transformative developments. Modern computer vision systems have shown remarkable success in medical imaging, autonomous driving, environmental perception, and remote sensing. However, these systems face two persistent challenges in real-world deployment: (i) their susceptibility to environmental variations and adversarial scenarios and (ii) their dependency on training datasets that may not sufficiently represent rare or degraded conditions.

This dissertation addresses these two limitations through independent yet complementary research directions. The first focuses on improving deep classification models' robustness and generalization capacity via correlation-penalized ensemble learning. The second introduces a physically grounded algorithm for simulating spatially inhomogeneous fog in outdoor scenes based on the discretized Radiative Transfer Equation (RTE) [1]. Both contributions aim to increase the reliability of vision systems under uncertainty and represent substantial methodological advancements in their respective areas.

The rise of convolutional neural networks (CNNs) has led to unprecedented accuracy in visual recognition tasks. However, individual CNNs are known to overfit training distributions and perform poorly under distributional shifts. Ensemble learning has emerged as a viable strategy to mitigate such brittleness by aggregating predictions from multiple base models. Crucially, the strength of an ensemble relies not only on the individual accuracy of its members but also on their diversity. When ensemble members make uncorrelated errors, the result-

ing model tends to be more robust, better calibrated, and more reliable in safety-critical scenarios.

Despite its theoretical promise, inducing diversity in deep ensembles remains a challenge. Traditional techniques—such as training with different seeds, using heterogeneous architectures, or applying data augmentations—often result in ensembles with limited functional diversity. Moreover, offline aggregation (e.g., averaging or majority voting) does not explicitly encourage de-correlated predictions during training.

To address this, we propose a novel online ensemble learning framework that directly integrates a Pearson correlation-based penalty into the loss function. This penalty term discourages ensemble members from producing strongly correlated incorrect predictions while allowing agreement on correct classifications. By embedding this regularization in a shared training pipeline, the proposed architecture jointly optimizes for accuracy and diversity. The resulting ensembles exhibit improved out-of-distribution robustness, better misclassification de-correlation, and reduced joint failure rates.

Empirical evaluation on four benchmark datasets - including natural scenes (CIFAR-10 [2], ISINI [3]) and clinical images (Diabetic Retinopathy [4, 5], ISIC skin lesions [6]) - confirms that our method consistently outperforms classical ensemble fusion strategies (e.g., majority voting, weighted averaging). It further demonstrates that moderate penalization (e.g.,  $\lambda \in [0.5, 1.0]$ ) yields the best balance between classification accuracy and output de-correlation. The method is computationally efficient, architecture-agnostic, and deployable in real-time systems, making it suitable for both academic and industrial applications.

Although ensemble diversity addresses model-side robustness, the second research line tackles data-side limitations. Training data in real-world computer vision is often insufficient to capture complex visual degradations, such as atmospheric scattering caused by fog. Traditional synthetic fog generation methods, including those based on Koschmieder’s law [7], assume homogeneity and isotropic scattering. These assumptions simplify computation but fail to reproduce realistic, depth-dependent, nonhomogeneous visual effects—thereby limiting the usefulness of such synthetic data for training robust perception models.

In response, we develop a physically grounded simulation technique that discretizes the RTE to model inhomogeneous, anisotropically scattering fog. The RTE captures how radiance propagates through participating media, accounting for absorption and scattering in all directions. Our method discretizes both spatial and angular domains and applies a recursive formulation that estimates radiance attenuation and in-scattering per pixel. The framework supports depth-aware attenuation, forward-biased scattering (via Henyey-Greenstein phase functions), and spatial variability in extinction coefficients.

To ensure tractability, we implement the algorithm using tensorized operations that allow efficient parallelization and gradient-based optimization. This enables integration into training pipelines for fog-aware computer vision tasks, while maintaining physical interpretability. The method is further enhanced by a calibration procedure using annotated foggy – cloudy image pairs, which allows the model to learn real-world fog distributions.

Experimental comparisons against Koschmieder model and GAN-based [8] fog generator show that our approach achieves superior realism and structure preservation, as measured by metrics such as Fréchet Inception Distance (FID) [9], Learned Perceptual Image Patch Similarity (LPIPS) [10], and Pearson correlation. These results suggest that physics-driven simulation can serve as a reliable foundation for data set enhancement, ultimately improving the robustness of the downstream model under visibility degradation.

While ensemble learning and fog simulation may appear distinct, they share a unifying goal: to increase the robustness of computer vision systems facing environmental uncertainty. The former addresses the variability in model behavior through diversity-aware optimization; the latter enhances the training data with realistic adverse conditions grounded in physical laws.

Together, these contributions form a cohesive response to a key challenge in applied AI: ensuring dependable operation in the wild. They enable models not only to generalize better across datasets but also to adapt to rare and challenging visual scenarios that are difficult to capture through manual annotation or standard augmentation techniques.

By combining methodological rigor with practical applicability, the research advances the state of the art in both algorithmic ensemble design and data synthesis. It contributes tools and insights that can benefit a wide range of applications, from medical diagnostics to autonomous navigation, and lays the groundwork for future integration of physically inspired modeling and learning-based systems.

## 2 Increasing CNN Ensemble Diversity via Correlation Penalty

CNNs have become the cornerstone of modern image classification due to their exceptional ability to extract hierarchical features from image data. However, despite their success, single CNNs often struggle with generalization, particularly when facing out-of-distribution inputs or noisy environments. Ensemble learning [11] provides a way to mitigate these limitations by combining multiple models to improve accuracy and robustness. The key to an effective ensemble lies not only in the individual strength of its members but in their diversity [12]—the extent to which they make different errors.

In this research, we proposed a novel ensemble learning framework that explicitly promotes diversity among CNN ensemble members through a Pearson correlation-based penalty integrated into the joint training loss. Unlike traditional ensemble approaches, which typically rely on offline aggregation of independently trained models, our method trains all ensemble members simultaneously in an online fashion, enabling the optimization process to consider both accuracy and output diversity.

Standard ensemble techniques such as majority voting [13, 14, 15] or averaging [16, 17, 18, 19, 20] often fail to fully exploit the potential of model ensembles due to the high correlation in the outputs of similarly trained networks. Classical methods to induce diversity—e.g., random initialization, data augmentation [14, 21], or varying architectures [16, 22, 23]—remain indirect and often insufficient.

To address this, we introduced a loss function of the form:

$\mathcal{L} = \mathcal{L}_{\text{CE}} + \lambda \mathcal{L}_{\text{corr}}$  where  $\mathcal{L}_{\text{CE}}$  is the categorical cross-entropy loss,  $\mathcal{L}_{\text{corr}}$  is a correlation-based penalty term, and  $\lambda$  is a hyperparameter controlling the trade-off between accuracy and diversity. The correlation penalty is computed using the Pearson correlation coefficient between the softmax outputs of ensemble members, and it is designed to penalize agreement on incorrect predictions while allowing agreement on correct classifications.

Each ensemble member is a CNN with its final fully connected layer replaced by a custom task-specific head. The outputs of these networks are passed through a shared trainable fusion layer, which aggregates predictions and applies a final softmax activation. During training, the gradient of the correlation penalty is backpropagated along with the standard cross-entropy loss. The penalty is computed across all model pairs in the ensemble, ensuring that models are not merely redundant copies of each other but instead learn diverse patterns from the data.

The Pearson correlation coefficient between output vectors  $h_i$  and  $h_j$  is given by:  $\rho(h_i, h_j) := \frac{\text{Cov}(h_i, h_j)}{\sigma_{h_i} \sigma_{h_j}}$  where  $\text{Cov}(h_i, h_j)$  denotes the covariance and  $\sigma$  represents the standard deviation. The penalty encourages  $\rho(h_i, h_j)$  to be low when both predictions are incorrect.

The proposed method was evaluated on four well-known image datasets: ISINI and CIFAR-10 (natural images), and Diabetic Retinopathy (DR) and ISIC (clinical images). Both homogeneous (e.g., multiple AlexNets [24] or VGG16s [25]) and heterogeneous ensembles (e.g., AlexNet + ResNet50 [26] + MobileNetV2 [27]) were tested. The results demonstrate the effectiveness of the proposed method in various scenarios:

- The method consistently outperformed classical ensemble strategies (majority voting, averaging, weighted voting) in terms of accuracy.
- Diversity metrics (double/triple miss counts) improved significantly with moderate values of  $\lambda$  (0.5 or 1.0).
- The method is robust across domains and architectures, showing especially strong gains in medical image classification tasks.
- The  $\lambda$  parameter offers explicit control over the diversity-accuracy trade-off.

**Thesis 2.1.** I created a new CNN ensemble framework in which the member networks are trained jointly through a loss function that integrates a Pearson correlation-based penalty. This formulation explicitly discourages redundant misclassifications, enhances functional diversity, and allows the ensemble to be optimized as a single end-to-end model. The regularization strength, controlled by the  $\lambda$  parameter, provides an explicit mechanism to balance classification accuracy and prediction diversity, enabling adaptation to domain- or task-specific requirements.

**Thesis 2.2.** The proposed online training approach outperforms classical ensemble strategies such as majority voting and averaging across all tested domains, demonstrating its superiority in promoting robust and complementary model behavior.

In conclusion, the proposed strategy for training marks a significant step forward in designing robust and generalizable CNN ensembles. By embedding a Pearson correlation

penalty into the learning process, the framework provides a direct, tunable mechanism to enforce functional diversity. Its empirical success across natural and clinical datasets supports its applicability in both academic research and safety-critical deployment scenarios.

The detailed experimental results and our findings have been published in [\[28\]](#)

### 3 Physically Based Fog Modeling in Inhomogeneous Media

Realistic simulation of environmental conditions is crucial for training and evaluating computer vision systems, particularly in safety-critical domains such as autonomous driving, aerial navigation, and surveillance. Among adverse weather conditions, fog represents a particularly challenging phenomenon, as it introduces complex light scattering and attenuation effects that significantly degrade visual information. While classical fog augmentation methods based on Koschmieder’s law have been widely used [29, 30], they are fundamentally limited by their simplifying assumptions, notably spatial homogeneity and isotropic scattering.

In contrast, data-driven methods such as CycleGAN [8] and StarGAN [31, 32] offer domain translation capabilities to generate fog effects. Although these generative adversarial networks can produce visually convincing results, they lack physical interpretability and controllability. Their outputs often suffer from style-based artifacts, geometric inconsistencies, and an inability to reflect physically meaningful fog behavior, as reported in [33, 34, 35]. Moreover, these models are inherently stochastic, making them less suitable for safety-critical applications where reproducibility and transparency are essential.

In this research, we proposed a novel fog synthesis framework grounded in the physics of light transport. Our approach is based on a discretized version of the Radiative Transfer Equation (RTE), a comprehensive model that governs the interaction of light with participating media. The proposed

method enables the simulation of fog with spatial inhomogeneity and anisotropic scattering effects, thereby addressing the shortcomings of traditional and AI based models.

The RTE formulation captures the attenuation of light along a path due to absorption and scattering, as well as the in-scattered radiance from other directions. To make this formulation tractable for image-space fog synthesis, we discretized both the spatial and angular domains. Each pixel ray is divided into depth steps, and angular integration is approximated using a sampled set of directions. To account for directional scattering, we employed the Henyey-Greenstein phase function. Our implementation models the full 3D geometry of fog distribution and light propagation within a single 2D image.

To overcome the computational challenges posed by this complex model, we developed a tensor-based implementation that reduces the overall computational cost of generating an  $n \times n$  foggy image from  $\mathcal{O}(n^5)$  to  $\mathcal{O}(n)$ . This acceleration makes the proposed framework practical for generating synthetic datasets. Moreover, our model supports differentiable rendering, enabling gradient-based optimization of physical parameters. This allows us to calibrate fog properties using real foggy-clear image pairs and adapt the synthesis process to varied atmospheric conditions.

We conducted extensive quantitative and qualitative evaluations comparing our method to both the analytical Koschmieder-based model and a recent generative adversarial network-based model. Using perceptual similarity metrics such as Fréchet Inception Distance (FID) and Learned Perceptual Image Patch Similarity (LPIPS), we demonstrated that our RTE-

based method generates foggy images with significantly improved realism and physical consistency. Additionally, the method preserves the geometric structure of the scene due to the use of depth maps, which are obtained using monocular estimation networks or stereo imaging.

The proposed framework bridges the gap between physical modeling and deep learning. It provides a robust tool for generating high-quality foggy imagery and lays the foundation for future work in physically grounded data augmentation, adversarial testing, and uncertainty modeling in perception systems.

**Thesis 3.1.** I proposed a new fog synthesis framework based on the discretized Radiative Transfer Equation (RTE), which operates in image space using monocular depth maps. By integrating the anisotropic Henyey–Greenstein phase function, the method accurately captures forward-scattering effects, enabling physically consistent and perceptually realistic fog generation that surpasses traditional isotropic formulations.

**Thesis 3.2.** The recursive numerical formulation of the discretized RTE reduces to the classical Koschmieder model under homogeneous and isotropic assumptions, thereby generalizing this widely used analytical law to inhomogeneous and anisotropic media.

**Thesis 3.3.** The proposed RTE-based simulation consistently outperforms both analytical and GAN-based fog synthesis approaches in terms of perceptual realism (LPIPS), distributional alignment (FID), and structural correlation. Furthermore, the recursive discretization was significantly optimized through a tensor-based implementation, which makes the method computationally much more efficient and scalable

in practice. The framework remains fully differentiable and efficiently implementable in deep learning environments, making it suitable for gradient-based parameter learning and large-scale synthetic dataset generation.

In conclusion, this research contributes to a physically grounded and differentiable fog generation algorithm that is capable of producing realistic fog effects in inhomogeneous media. Its integration of physical optics, numerical methods, and efficient GPU implementation enables its use in training and validating vision systems under diverse and challenging visual conditions. The method is not only theoretically sound but also practically effective, paving the way for further exploration in physically informed data augmentation for robust AI.

The detailed experimental results and our findings have been published in [\[36\]](#)

## References

- [1] S. Chandrasekhar. *Radiative Transfer*. Dover Books on Physics. Dover Publications, 2013.
- [2] Alex Krizhevsky, Vinod Nair, and Geoffrey Hinton. Cifar-10 - Canadian Institute for Advanced Research. *MIT*, 2009. from <http://www.cs.toronto.edu/~kriz/cifar.html>.
- [3] Roy Prasun, Ghosh Subhankar, Bhattacharya Saumik, and Pal Umapada. Effects of degradations on deep neural network architectures. *ArXiv*, 07 2018. from <https://arxiv.org/pdf/1807.10108.pdf>.
- [4] Kaggle. Diabetic Retinopathy Detection, 2015. from <https://www.kaggle.com/c/diabetic-retinopathy-detection>.
- [5] Etienne Decencière, Xiwei Zhang, Guy Cazuguel, Bruno Lay, Béatrice Cochener, Caroline Trone, Philippe Gain, Richard Ordonez, Pascale Massin, Ali Erginay, Béatrice Charton, and Jean-Claude Klein. Feedback on a publicly distributed database: the messidor database. *Image Analysis & Stereology*, 33(3):231–234, Aug 2014.
- [6] Noel C. F. Codella, David Gutman, M. Emre Celebi, Brian Helba, Michael A. Marchetti, Stephen W. Dusza, Aadi Kallou, Konstantinos Liopyris, Nabin Mishra, Harald Kittler, and Allan Halpern. Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic). In

- 2018 IEEE 15th International Symposium on Biomedical Imaging*, pages 168–172, 2018.
- [7] Hans Israël and Fritz Kasten. *KOSCHMIEDERs Theorie der horizontalen Sichtweite*, pages 7–10. VS Verlag für Sozialwissenschaften, Wiesbaden, 1959.
  - [8] Zaher Ghais. Simulating weather conditions on digital images. Master’s thesis, University of Debrecen, 2020.
  - [9] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS’17*, page 6629–6640, Red Hook, NY, USA, 2017. Curran Associates Inc.
  - [10] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.
  - [11] L.K. Hansen and Peter Salamon. Neural network ensembles. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 12:993 – 1001, 11 1990.
  - [12] Michael Perrone and Leon Cooper. When networks disagree: Ensemble methods for hybrid neural networks. *Neural networks for speech and image processing*, 08 1993.
  - [13] Yushi Chen, Ying Wang, Yanfeng Gu, Xin He, Pedram Ghamisi, and Xiuping Jia. Deep learning ensemble for

- hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(6):1882–1897, 2019.
- [14] Rodrigo Minetto, Maurício Pamplona Segundo, and Sudeep Sarkar. Hydra: An ensemble of convolutional neural networks for geospatial land classification. *IEEE Transactions on Geoscience and Remote Sensing*, 57(9):6530–6541, 2019.
- [15] Shuxian Dong, Wei Feng, Yinghui Quan, Gabriel Dauphin, Lianru Gao, and Mengdao Xing. Deep ensemble cnn method based on sample expansion for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–15, 2022.
- [16] Ishtiaque Ahmed Khan, Asaduzzaman Sajeeb, and Shaikh Anowarul Fattah. An automatic ocular disease detection scheme from enhanced fundus images based on ensembling deep cnn networks. In *2020 11th International Conference on Electrical and Computer Engineering*, pages 491–494, 2020.
- [17] Parham M. Kebria, Abbas Khosravi, Syed Moshfeq Salaken, and Saeid Nahavandi. Deep imitation learning for autonomous vehicles based on convolutional neural networks. *IEEE/CAA Journal of Automatica Sinica*, 7(1):82–95, 2020.
- [18] Wenmei Li, Haiyan Liu, Yu Wang, Zhuangzhuang Li, Yan Jia, and Guan Gui. Deep learning-based classification methods for remote sensing images in urban built-up areas. *IEEE Access*, 7:36274–36284, 2019.

- [19] Xiangrong Zhang, Wenkang Ma, Chen Li, Jie Wu, Xu Tang, and Licheng Jiao. Fully convolutional network-based ensemble method for road extraction from aerial images. *IEEE Geoscience and Remote Sensing Letters*, 17(10):1777–1781, 2020.
- [20] Baihua Zhang, Shouliang Qi, Patrice Monkam, Chen Li, Fan Yang, Yu-Dong Yao, and Wei Qian. Ensemble learners of multiple deep cnns for pulmonary nodules classification using ct images. *IEEE Access*, 7:110358–110371, 2019.
- [21] Andrey Kuehlkamp, Allan Pinto, Anderson Rocha, Kevin W. Bowyer, and Adam Czajka. Ensemble of multi-view learning classifiers for cross-domain iris presentation attack detection. *IEEE Transactions on Information Forensics and Security*, 14(6):1419–1431, 2019.
- [22] Najd Alosaimi and Haikel Alhichri. Fusion of cnn ensemble for remote sensing scene classification. In *2020 3rd International Conference on Computer Applications Information Security (ICCAIS)*, pages 1–6, 2020.
- [23] Ali Yazdizadeh, Zachary Patterson, and Bilal Farooq. Ensemble convolutional neural networks for mode inference in smartphone travel survey. *IEEE Transactions on Intelligent Transportation Systems*, 21(6):2232–2239, 2020.
- [24] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C.J. Burges, L. Bottou, and K.Q. Weinberger, editors, *Advances in Neural Informa-*

- tion Processing Systems*, volume 25. Curran Associates, Inc., 2012.
- [25] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition, 2015.
  - [26] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.
  - [27] Andrew Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv*, 04 2017.
  - [28] Balazs Harangi, Agnes Baran, Marcell Beregi-Kovacs, and Andras Hajdu. Composing diverse ensembles of convolutional neural networks by penalization. *Mathematics*, 11(23), 2023. (Q2, IF: 2.3, SJR: 0.498).
  - [29] Jean-Philippe Tarel and Nicolas Hautière. Fast visibility restoration from a single color or gray level image. In *2009 IEEE 12th International Conference on Computer Vision*, pages 2201–2208, 2009.
  - [30] Jean-Philippe Tarel, Nicolas Hautière, Aurélien Cord, Dominique Gruyer, and Houssam Halmaoui. Improved visibility of road scene images under heterogeneous fog. In *2010 IEEE Intelligent Vehicles Symposium*, pages 478–485, 2010.

- [31] Blessing Agyei Kyem, Joshua Kofi Asamoah, Ying Huang, and Armstrong Aboah. Weather-adaptive synthetic data generation for enhanced power line inspection using star-gan. *IEEE Access*, 12:193882–193901, 2024.
- [32] Yating Lin, Yidong Li, Haidong Cui, and Zheng Feng. Weagan:generative adversarial network for weather translation of image among multi-domain. In *2019 6th International Conference on Behavioral, Economic and Socio-Cultural Computing (BESC)*, pages 1–5, 2019.
- [33] Fayçal Abbas and Med Babahenini. Forest fog rendering using generative adversarial networks. *The Visual Computer*, 39:1–10, 01 2022.
- [34] Rui Gong, Dengxin Dai, Yuhua Chen, Wen Li, Danda Paudel, and Luc Gool. Analogical image translation for fog generation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35:1433–1441, 05 2021.
- [35] Hanqing Zhang, Qitao Dan, and Lingfeng Wang. From point to surface: Realistic and perceptually-plausible hazy image generation with glow-diffusion. In *Pattern Recognition and Computer Vision: 7th Chinese Conference, PRCV 2024, Urumqi, China, October 18–20, 2024, Proceedings, Part X*, page 89–102, Berlin, Heidelberg, 2024. Springer-Verlag.
- [36] Marcell Beregi-Kovacs, Balazs Harangi, Andras Hajdu, and Gyorgy Gat. Generation of synthetic non-homogeneous fog by discretized radiative transfer equation. *Journal of Imaging*, 11(6), 2025. (Q1, IF: 3.3, SJR: 0.662).



Registry number: DEENK/422/2025.PL  
Subject: PhD Publication List

Candidate: Marcell Beregi-Kovács

Doctoral School: Doctoral School of Mathematical and Computational Sciences

MTMT ID: 10083987

## List of publications related to the dissertation

### Foreign language scientific articles in international journals (2)

1. **Beregi-Kovács, M.**, Harangi, B., Hajdu, A., Gát, G.: Generation of Synthetic Non-Homogeneous Fog by Discretized Radiative Transfer Equation.  
*J. Imaging.* 11 (6), 1-22, 2025. EISSN: 2313-433X.  
DOI: <http://dx.doi.org/10.3390/jimaging11060196>  
IF: 3.3 (2024)
2. Harangi, B., Baran, Á., **Beregi-Kovács, M.**, Hajdu, A.: Composing Diverse Ensembles of Convolutional Neural Networks by Penalization.  
*Mathematics.* 11 (23), 1-19, 2023. EISSN: 2227-7390.  
DOI: <http://dx.doi.org/10.3390/math11234730>  
IF: 2.3

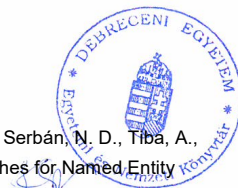
## List of other publications

### Foreign language scientific articles in Hungarian journals (1)

3. Bogacsovcics, G., Hajdu, A., Lakatos, R., **Beregi-Kovács, M.**, Tiba, A., Tomán, H.: Replacing the SIR epidemic model with a neural network and training it further to increase prediction accuracy.  
*Ann. Math. Inform.* 53, 73-91, 2021. ISSN: 1787-5021.  
DOI: <http://dx.doi.org/10.33039/ami.2021.02.003>

### Foreign language conference proceedings (2)

4. Bogacsovcics, G., Harangi, B., **Beregi-Kovács, M.**, Kupás, D., Lakatos, R., Serbán, N. D., Tiba, A., Tóth, J.: Assessing Conventional and Deep Learning-Based Approaches for Named Entity Recognition in Unstructured Hungarian Medical Reports.  
In: 2024 IEEE 22nd World Symposium on Applied Machine Intelligence and Informatics (SAMI). Ed.: Kovács Levente, Liberios Vokorokos, IEEE, Piscataway, 77-82, 2024. ISBN: 9798350317206





5. **Beregi-Kovács, M.**, Baran, Á., Hajdu, A.: Efficient Learning of Model Weights via Changing Features During Training.

In: 2020 IEEE 24th International Conference on Intelligent Engineering Systems (INES), Institute of Electrical and Electronics Engineers (IEEE), Piscataway (NJ), 43-48, 2020. ISBN: 9781728110592

**Total IF of journals (all publications): 5,6**

**Total IF of journals (publications related to the dissertation): 5,6**

The Candidate's publication data submitted to the Tudóstér have been validated by DEENK on the basis of the Journal Citation Report (Impact Factor) database.

24 June, 2025

