

Super-rezoluție video

Victor-Eduard Robu

Rezumat

Problema super-rezoluției se referă la scalarea imaginilor și recuperarea detaliilor, pe baza relațiilor dintre pixeli vecini. Super-rezoluția video aduce, în plus, informațiile datorate pixelilor din cadrele învecinate unui cadru central. Problema nu are soluție unică, deoarece se dorește obținerea mai multor pixeli, plecând de la unul singur, fără utilizarea altor informații suplimentare. De aceea, au fost propuse multe soluții de-a lungul timpului, care se bazează pe impunerea unor constrângeri între pixeli învecinați pentru a obține o soluție unică. Cele mai recente și performante metode folosesc rețele neuronale, sisteme ce au luat avânt în domeniul învățării automate în ultimii ani, datorită descoperirilor recente și a creșterii puterii de calcul.

Lucrarea de față își propune implementarea unor soluții pentru super-rezoluția secvențelor video în framework-ul Tensorflow, bazate pe două modele existente, ESPCN și FRVSR, ambele fiind bazate pe rețele neuronale. ESPCN este o arhitectură simplă, al cărui scop este super-rezoluția imaginilor, ce conține o rețea cu straturi de convoluție. Prin comparație, FRVSR este o arhitectură mai recentă și mai complexă. Își propune super-rezoluția secvențelor video, prin intermediul unui model recurent ce conține mai multe componente, dintre care una este responsabilă cu găsirea corespondențelor pixelilor dintre cadre, date prin flux optic, și o alta realizează operația de scalare propriu-zisă, folosind un cadru și estimarea de rezoluție înaltă a cadrului anterior. În urma evaluării, se constată faptul că arhitectura FRVSR este capabilă să folosească informații din cadrele învecinate, depășind performanțele rețelei ESPCN. Prin evaluarea modelului FRVSR, folosind mai multe modalități de calcul al fluxului optic, se determină că modulul responsabil de calculare a fluxului optic, format dintr-o rețea neuronală, nu este capabil să dea rezultate utile, motiv pentru care este înlocuit cu unul din algoritmi folosiți la evaluare, numit DIS. Validarea modelelor antrenate se realizează prin intermediul funcționalităților oferite de Tensorflow, dar și a unor rapoarte HTML ce conțin rezultatele antrenării pentru fiecare secvență, cadru și metodă testată.

Pentru antrenarea și evaluarea modelelor se folosește setul de date MPI-Sintel, bazat pe animația *open-source* Sintel, deoarece a fost proiectat pentru evaluarea metodelor de calcul al fluxurilor optice, motiv pentru care conține fluxul optic exact pentru scenele conținute.

Pentru expunerea modelelor antrenate, s-a implementat și o interfață grafică, folosind framework-ul Qt. Interfața permite selectarea unui videoclip și a unor metode de scalare, în vederea producerii unui videoclip care să le compare. Metodele sunt rețelele antrenate, dar și interpolările bicubică și biliniară.