

Automatische Unterscheidung von Verhaltensmustern bei Schweinen auf der Basis von Anomalieerkennung durch ein neuronales Konvolutionsnetzwerk

Martin Wutke¹, Mehmet Gültas^{1,3}, Imke Traulsen² und Armin O. Schmitt^{1,3}

Abstract: Die automatisierte Erfassung und Klassifizierung spezifischer Verhaltensmuster von Hausschweinen ermöglicht die Untersuchung unterschiedlicher Einflussfaktoren in den Haltungsbedingungen. Vor allem die Analyse von Videoaufnahmen von Tieren stellt bestehende Ansätze vor Herausforderungen, da die beobachtbaren Verhaltensmuster keiner bestimmaren Verteilung zu folgen scheinen. Die präsentierte Methode verwendet einen Machine-Learning-Algorithmus, um das Aktivitätsniveau verschiedener Schweinegruppen auf Basis von Videoaufnahmen zu bestimmen. In einem ersten Schritt wird ein neuronales Netzwerk darauf trainiert, Anomalien in Form von unerwarteten Aktivitäten in den Videodateien zu detektieren. Anhand der erzielten Ergebnisse wird in einem zweiten Schritt ein Klassifizierungsalgorithmus entwickelt, wodurch ein standardisierter Vergleich unterschiedlicher Videosequenzen ermöglicht wird.

Keywords: Verhaltensbestimmung; Maschinelles Lernen, Videoanalyse

1 Einleitung und Problemstellung

Die Förderung des Tierwohls im Rahmen der Nutztierhaltung steht seit Jahren verstärkt im Fokus des öffentlichen Interesses. Vor allem Schweine sind sozial lebende Tiere und die Gruppenhaltung gilt als anzustrebende Haltungsform [KT11]. Allerdings kann es hierbei zum Auftreten aggressiver Verhaltensmuster kommen, welche sich oftmals im Zusammenhang mit einem erhöhten Bewegungs- und Aktivitätsniveau zeigen [OC14]. Durch die Identifikation spezifischer Bewegungsmuster kann das Tierverhalten eingehender analysiert und Tätertiere können leichter identifiziert werden, wodurch langfristig agonistische Verhaltensweisen herausgezüchtet werden können, was wiederum zu einer Verbesserung des Tierwohls beiträgt.

Allerdings stellen sowohl die sofortige Identitätsbestimmung als auch die nachträgliche Identifizierung mittels Videoauswertung eine hohe finanzielle und zeitliche Belastung dar.

1 Georg-August-Universität Göttingen, Arbeitsgruppe Züchtungsinformatik, Margarethe von Wrangell-Weg 7, 37075 Göttingen, Deutschland, martin.wutke@stud.uni-goettingen.de; gultas@informatik.uni-goettingen.de; armin.schmitt@uni-goettingen.de

2 Georg-August-Universität Göttingen, Arbeitsgruppe Systeme der Nutztierhaltung, Albrecht-Thaer-Weg 3, 37075 Göttingen, Deutschland, imke.traulsen@uni-goettingen.de

3 Center for Integrated Breeding Research (CiBreed), Albrecht-Thaer-Weg 3, Georg-August University, 37075, Göttingen, Deutschland

Ziel dieser Arbeit war die Implementierung eines neuronalen Netzwerks zur automatisierten Verhaltensbestimmung von Hausschweinen auf Basis von Anomaliedetektion in Videoaufnahmen. Das trainierte Modell wird danach dazu verwendet, 30-minütige Videosequenzen zu analysieren. Basierend auf den Ergebnissen dieses Analyseschritts wird ein Klassifikationsmodell verwendet, um die Intensität des Aktivitätsniveaus zu bewerten und die Vergleichbarkeit zwischen verschiedenen Videos zu erhöhen.

2 Methodik

Zur Analyse der Videoaufnahmen wird ein *unsupervised Deep Learning-Algorithmus* in Form eines neuronalen Konvolutionsnetzwerkes (convolutional neural network, CNN) in einer Autoencoder-Struktur verwendet. Im Gegensatz zum derzeit dominierenden *supervised*-Ansatz werden bei einem *unsupervised*-Ansatz keine annotierten Daten für das Modelltraining benötigt [GBC16]. Als primäre Datenquelle werden extrahierte Schwarz-Weiß-Bilder in einer Auflösung von 400 x 600 Pixel einzelner Videoaufnahmen aus sechs verschiedenen Schweinebuchten verwendet.

Die Kernidee des implementierten Netzwerks besteht darin, die charakteristischen datengenerierenden Merkmale einer Sequenz von Einzelbildern (Frames) zu extrahieren und darauf basierend eine Vorhersage für das nachfolgende Bild zu erstellen. Hierbei werden die bildbeschreibenden Merkmale durch eine Autoencoder-Struktur, welche sich aus einem Encoder-Submodell und einem Decoder-Submodell zusammensetzt, extrahiert. Abbildung 1 zeigt diesbezüglich die verwendete Netzwerkstruktur und veranschaulicht die Systematik der Bildvorhersage.

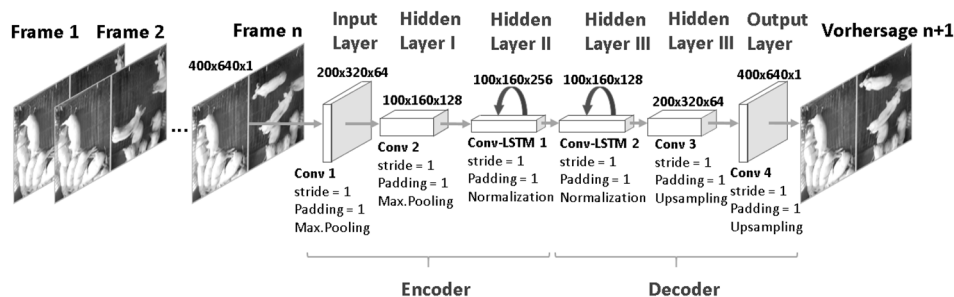


Abb. 1: Das neuronale Netzwerk weist eine Tiefe von sechs 2D-Konvolutionsschichten (Hidden Layers) auf. Für jede Konvolutionsoperation wird ein Stride- und Paddingparameter von eins sowie eine Max.Pooling-Operation verwendet, wodurch die Dimension der Eingangsdaten schichtweise verändert wird. Als Netzwerkoutput wird eine Vorhersage für das nachfolgende Bild erstellt und mit dem tatsächlichen Bild verglichen.

Für das Modelltraining werden ausschließlich Videosequenzen aus Ruhephasen verwendet, wodurch die Vorhersagegenauigkeit des trainierten Netzwerks für Videosequenzen mit niedrigen Aktivitätsniveaus erhöht wird. Videosequenzen mit hohen Bewegungs- und Aktivitätsniveaus werden hingegen verzerrt vorhergesagt. Durch einen paarweisen Vergleich der vorhergesagten Videoframes mit den tatsächlichen Frames kann die Güte der Vorhersage anhand des mittleren quadratischen Fehlers quantifiziert werden, wobei Videosequenzen mit hohen Rekonstruktionsfehlern und daher hohem Aktivitätsniveau als Anomalie eingestuft werden.

Aufbauend auf den Ergebnissen der Netzwerkanalyse wird auf Basis des mittleren quadratischen Rekonstruktionsfehlers ein Klassifikationsalgorithmus trainiert, wodurch unterschiedliche Videosequenzen in einem normierten Wertebereich von [1:100] verglichen werden können. Die untere Grenze des Wertebereichs entspricht hierbei dem niedrigsten Aktivitätsniveau und die Obergrenze 100 dem höchsten Aktivitätsniveau. Die Videoklassifikation ist beispielhaft für zwei Testvideos in Abbildung 2 dargestellt.

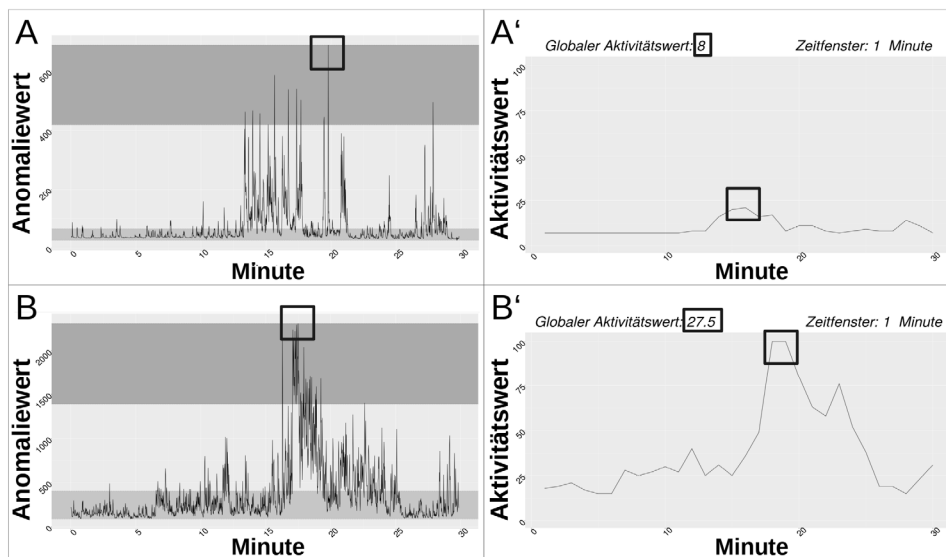


Abb. 2: (A) und (B) zeigen zwei vom Netzwerk analysierte Videos. Die unterschiedlichen Bereiche entsprechen unterschiedlichen Aktivitätsniveaus, wobei der untere Bereich für ein geringes Aktivitätsniveau und der obere Bereich für ein hohes Aktivitätsniveau steht. (A') und (B') zeigen die Ergebnisse des Klassifikationsalgorithmus. Im dargestellten Beispiel wird Video (A) mit einem durchschnittlichen globalen Aktivitätswert von 8 und Video (B) mit einem Wert von 27,5 eingestuft (Rahmen).

Die in Abbildung 2 dargestellten Videosequenzen weisen unterschiedliche Aktivitätsniveaus auf, was lediglich im Vergleich anhand der variierenden Skalenbereiche zu erkennen ist. Video (A) enthält überwiegend Phasen mit geringen Aktivitätsleveln, wohingegen Video (B) Phasen höherer Aktivität enthält. Durch die Klassifizierung der

Videsequenzen mittels des vorab trainierten Klassifikationsalgorithmus wird jeder Videosequenz ein globaler durchschnittlicher Aktivitätswert basierend auf dem individuellen Aktivitätsverhalten der Tiere zugeordnet, wodurch erste Erkenntnisse über das Tierverhalten gewonnen werden können. Darüber hinaus wird ein standardisierter Aktivitätsverlauf für ein vorab definiertes Zeitfenster berechnet, was die Identifikation, Filterung und den Vergleich von Phasen höherer Aktivität ermöglicht.

Für den Modelltest werden 288 nicht-annotierte Videos verwendet. Der Evaluierungsprozess ist zweigeteilt, wobei im ersten Schritt jedes Video einzeln ausgewertet wird. Um die Vergleichbarkeit der Videos untereinander zu erhöhen und um das Aktivitätsverhalten zu unterschiedlichen Zeitpunkten analysieren zu können, werden die Ergebnisse der Einzelauswertungen in einem zweiten Evaluierungsschritt verdichtet, indem die quadrierten Rekonstruktionsfehler für jede Videosequenz kumuliert werden.

3 Ergebnisse

Das Ergebnis des ersten Evaluierungsschrittes ist beispielhaft für ein Testvideo in Abbildung 3 aufgeführt.

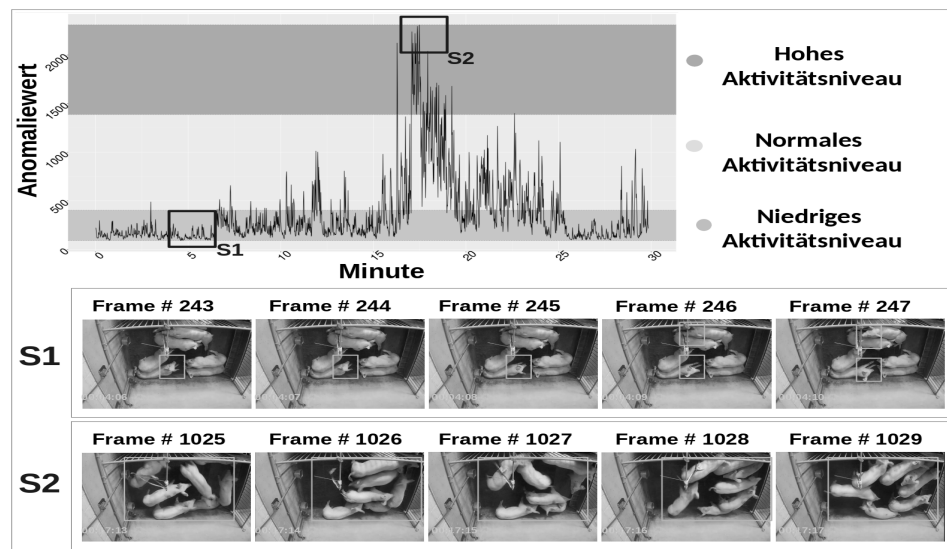


Abb. 3: Das 30-minütige Video weist zu Beginn eine Phase niedriger Aktivität auf (S1), welche durch leichte Kopfbewegungen einiger Tiere verursacht wird. In Minute 17 (S2) verschlechtert sich die Vorhersagequalität des Netzwerkes, was auf verstärkte Aktivität der Tiere hinweist. Bei Sichtung der entsprechenden Videostelle zeigt sich, dass die erhöhten Rekonstruktionsfehler durch starke Bewegungen aller Tiere hervorgerufen werden.

Für den zweiten Evaluierungsschritt werden die Ergebnisse der Einzelanalyse verdichtet. Das Ergebnis für alle 288 Videos ist in Abbildung 4 dargestellt.

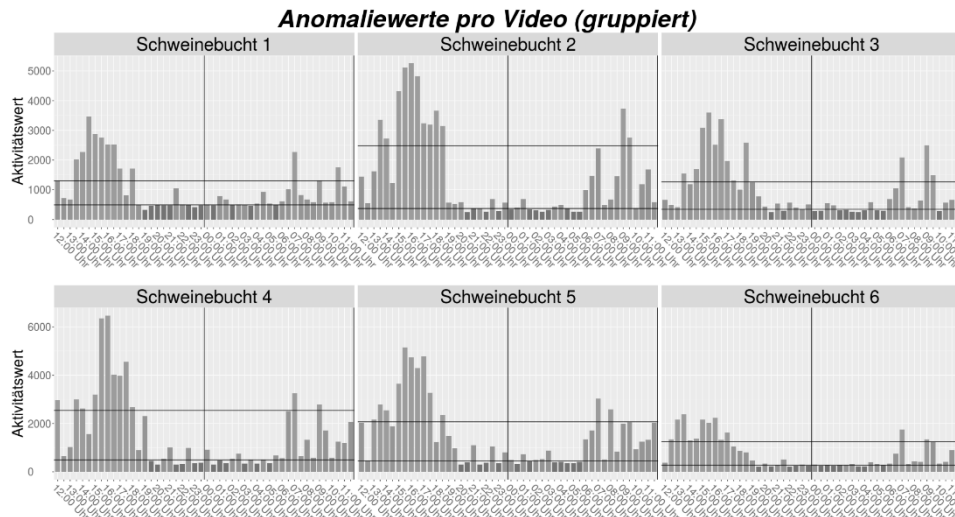


Abb. 4: Modellierung des Aktivitätsverlaufs unter Berücksichtigung gruppenspezifischer Effekte.

Abbildung 4 zeigt die Auswertung der sechs unterschiedlichen Schweinebuchten in einer Zeitspanne von 24 Stunden mit Beginn der Videoaufzeichnung um 12 Uhr. Jede Gruppe enthält 48 Videosequenzen. Im Ergebnis zeigt sich zunächst ein gruppenkonsistenter Anstieg des Aktivitätsniveaus der Tiere zu Beginn der Aufzeichnung. Mit Eintreten der Nacht fällt das Aktivitätsniveau erwartungsgemäß ab, was durch den natürlichen Schlafrhythmus der Tiere erklärt werden kann. Zum Ende der Aufzeichnung steigt das Aktivitätsniveau zwischen 6 Uhr und 11 Uhr wieder an.

Wie in Abbildung 4 zu erkennen ist, kommt es bei der Rekonstruktion von Videos unterschiedlicher Schweinebuchten zu einem gruppenspezifischen Effekt, wodurch ein Vergleich von Videosequenzen nur innerhalb einer Gruppe gegeben ist. Um diese Limitierung zu überwinden und um das Ergebnis einer einzelnen Videoanalyse besser einstufen zu können, wurden alle analysierten Videosequenzen durch einen zuvor angelegten Klassifikationsalgorithmus bewertet. Hierbei wurden die Rekonstruktionsfehler aller Sequenzen einer Gruppe gemeinsam betrachtet, um ein standardisiertes Klassensystem bestehend aus 100 Aktivitätsklassen zu erstellen. Das Ergebnis ist beispielhaft in Abbildung 2 dargestellt. Für jede Videosequenz kann auf diese Weise ein globaler Aktivitätswert bestimmt werden, wodurch ohne manuelle Auswertung erste Informationen über das Tierverhalten erlangt werden können. Darüber hinaus wird der Aktivitätsverlauf durch ein frei wählbares Zeitfenster approximiert. Da diese Aktivitätsstandardisierung für alle Videogruppen separat vorgenommen wird, wird zudem die Vergleichbarkeit gruppenübergreifend erhöht.

4 Fazit

Die Verwendung von Videoanalysen im Rahmen der Nutztierhaltung ermöglicht eine effizientere Untersuchung des Tierverhaltens. Der in dieser Arbeit vorgestellte Ansatz der automatisierten Analyse von Videodaten und der Erkennung unterschiedlicher Verhaltensmuster kann in einer Weiterentwicklung einen Beitrag zur Lösung der Problematik der Identifikation aggressiver Einzeltiere und damit zu einer langfristigen Steigerung des Tierwohls leisten. Darüber hinaus wird der zeitliche Aufwand der manuellen Auswertung reduziert, was eine Kosten- und Fehlerreduktion mit sich führt. Die Nutzung und das Potenzial neuronaler Netzwerke steht in diesem Zusammenhang noch in der Entwicklungsphase und zukünftige Netzwerkalgorithmen könnten beispielsweise Einzeltiere lokalisieren oder spezifische Verhaltenssituationen erfolgreich detektieren.

Literaturverzeichnis

- [GBC16] Goodfellow, I.; Bengio, Y.; Courville, A.: Deep Learning (Adaptive Computation and Machine Learning series). The MIT Press, 2016.
- [KT11] KTBL: Ökologische Schweinehaltung – zukunftsweisende Haltungsverfahren. Kuratorium für Technik und Bauwesen in der Landwirtschaft (Darmstadt), 2011.
- [Oc14] Oczak, M.; Viazzi, S.; Ismayilova, G.; Sonoda, L. T.; Roulston, N.; Fels, M.; Bahr, C.; Hartung, J.; Guarino, M.; Berckmans, D. et al.: Classification of aggressive behaviour in pigs by activity index and multilayer feed forward neural network. Biosystems Engineering 119, S. 89-97, 2014.