

LE DEEP LEARNING POUR L'ESTIMATION DE LA DISTRIBUTION EN TAILLE DE PARTICULES DE TiO_2 À PARTIR D'IMAGES EN MICROSCOPIE ÉLECTRONIQUE À BALAYAGE

Loïc Coquelin¹ & Paul Monchot² & Nicolas Fischer³ & Nicolas Feltin⁴ & Alexandra Delvallée⁵

¹ *Laboratoire National de Métrologie et d'Essais, France, Loic.Coquelin@lne.fr*

² *Laboratoire National de Métrologie et d'Essais, France, Paul.Monchot@lne.fr*

³ *Laboratoire National de Métrologie et d'Essais, France, Nicolas.Fischer@lne.fr*

⁴ *Laboratoire National de Métrologie et d'Essais, France, Nicolas.Feltin@lne.fr*

⁵ *Laboratoire National de Métrologie et d'Essais, France, Alexandra.Delvallee@lne.fr*

Résumé. Afin de détecter et caractériser les particules de TiO_2 présentes dans les produits alimentaires et cosmétiques (Weir *et al.* (2012), Hwang *et al.* (2019)), le microscope électronique à balayage (MEB) est généralement l'instrument privilégié avec des images haute résolution permettant de mesurer la taille, la forme ou l'état d'agrégation des particules. Pour comprendre les propriétés des matériaux à l'échelle nanométrique, il s'agit d'estimer précisément la distribution en taille des particules et pour se faire, il faut détecter et segmenter chaque particule individuelle dans l'image et estimer leur état d'agrégation. Nous proposons d'utiliser le Mask-RCNN développé par He *et al.* (2017) pour la tâche de segmentation automatique des particules avec un entraînement couplant les techniques de transfert d'apprentissage et d'augmentation de données compte-tenu du nombre limité d'images annotées (courant dans l'industrie). Pour estimer le statut d'agrégation de chaque particule, nous utilisons un réseau de neurone convolutionnel (VGG16). Cette chaîne de traitement automatisée permet une bonne estimation de la distribution en taille des particules de TiO_2 avec 96 % des mesures sur le jeu de test présentant moins de 5 % d'erreur sur l'estimation du diamètre de surface équivalent.

Mots-clés. Microscopie Électronique à Balayage, Particules de TiO_2 , Réseaux de neurones, Segmentation d'instances, Classification ...

Abstract. In order to detect and characterize TiO_2 particles present in food and cosmetic products (Weir *et al.* (2012), Hwang *et al.* (2019)), the scanning electron microscope (SEM) is generally the preferred instrument with high-resolution images enabling measurement of particle size, shape or aggregation state. To understand the properties of nanoscale materials, we need to accurately estimate the particle size distribution. To do this, we need to detect and segment each individual particle in the image, and estimate their state of aggregation. We propose to use the Mask-RCNN developed by He *et al.* (2017) for the automatic particle segmentation task with training coupling transfer learning and data augmentation techniques given the limited number of annotated images (common in the industry). To estimate the aggregation status of each particle, we use a convolutional neural network (VGG16). This automated pipeline provides a good estimate of the size distribution of TiO_2 particles, with 96% of measurements on the test set showing less than 5% error in equivalent surface diameter estimation.

Keywords. Scanning Electron Microscopy, TiO_2 particles, Neural Networks, Instance Segmentation, Classification . . .

1 Introduction

A partir d'images acquises au microscope électronique à balayage (MEB), comme l'exemple présenté en figure 1, la chaîne de traitements aboutissant à l'estimation de la distribution en taille des particules de dioxyde de titane (TiO_2)¹ comporte plusieurs étapes : la détection et la segmentation de chaque particule présente dans l'image, la classification de l'état d'agrégation de chaque particule et enfin le calcul du diamètre équivalent (diamètre de Feret ou diamètre de surface équivalent, ...). On peut également procéder à la complétion des particules partiellement visibles comme proposé par Coquelin et *al.* (2019). Bien que chaque étape doit faire l'objet d'une attention particulière, l'étape clé reste évidemment la segmentation des particules, en effet, les autres étapes dépendent directement des performances initiales de segmentation. Les sections 2 et 3 présentent les développements spécifiques réalisés pour ce cas d'étude respectivement pour les tâches de segmentation et de classification pour finir en section 4 par la présentation des performances obtenues sur les différents jeux de données de test.

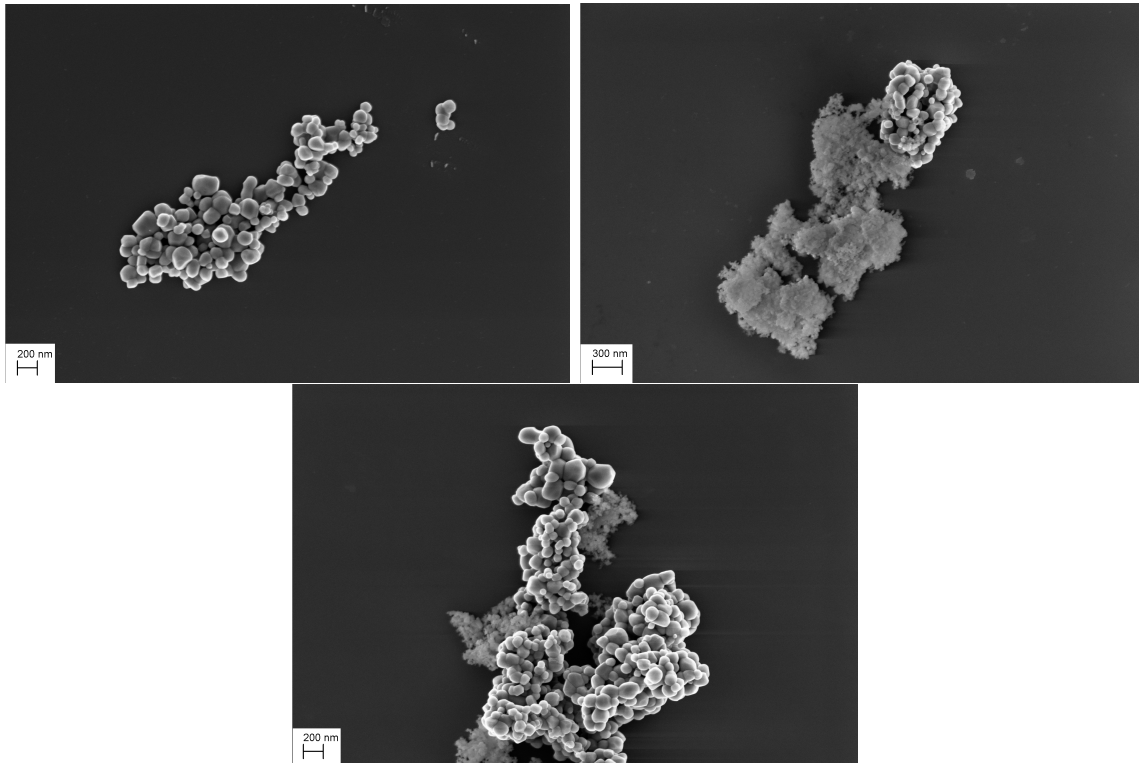


Figure 1: Exemples d'images de particules de TiO_2 obtenues au MEB présentant des agrégats typiques avec pour deux images la présence à l'arrière plan de particules de SiO_2 .

¹la distribution en taille est obtenue en estimant le diamètre équivalent de chaque particule de l'image

2 Segmentation des particules de TiO_2

La base de données métier La base de données est constituée de 77 images en niveaux de gris de taille 2048×1536 segmentées manuellement par des experts en nanométrie pour un total de 5947 particules de TiO_2 . La segmentation d'une particule par un annotateur correspond à un masque binaire.

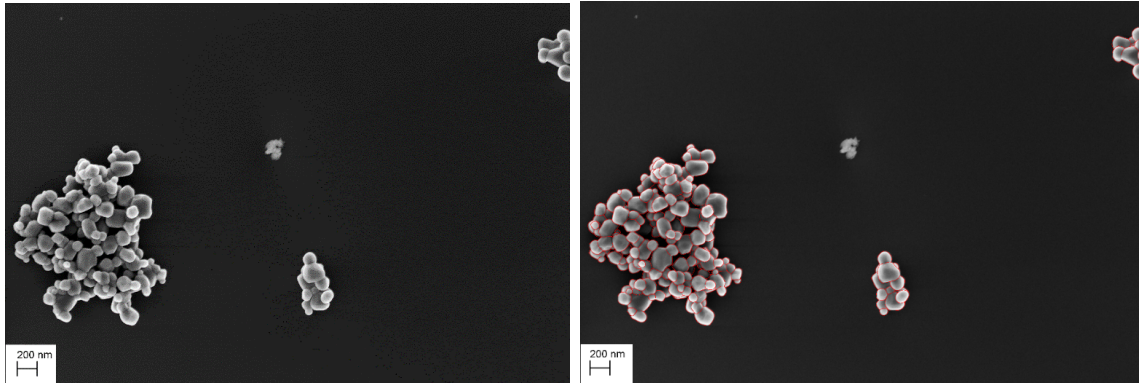


Figure 2: Exemple d'image de particules de TiO_2 mesurées au MEB et sa version segmentée.

Le réseau de neurones sélectionné L'algorithme Mask-RCNN développé par He *et al.* (2017) utilise trois réseaux de neurones : le premier (backbone network) est utilisé pour extraire des caractéristiques de l'image en entrée; le deuxième (region proposal network ou RPN) est utilisé pour générer des régions candidates contenant des objets d'intérêt²; le dernier réseau (mask head) prend en entrée les régions proposées et génère des masques binaires pour chaque objet détecté dans l'image. Nous utilisons un réseau pyramidal (FPN, Lin *et al.* (2017)) comme backbone afin de capturer des informations détaillées à des échelles fines ainsi que des informations contextuelles à des échelles plus larges dans l'image. Pour fusionner les caractéristiques extraites provenant de différentes résolutions spatiales, des connexions résiduelles sont employées (ResNet-50-FPN). Enfin, afin de s'assurer un haut niveau de résolution pour la génération de masques, des couches supplémentaires ont été ajoutées à l'architecture initialement proposée pour la génération de masques.

Comme souvent lorsque l'on souhaite utiliser les récents développements dans la communauté de l'apprentissage statistique (Machine Learning ou Deep Learning) dans l'industrie, la question du nombre de données disponibles peut s'avérer être un verrou à l'utilisation. Conscient du nombre limité de données à disposition, nous avons recours aux techniques dites de l'augmentation de données et du transfert d'apprentissage.

²il produit des boîtes englobantes ou "ancres" avec des scores associés en parcourant les cartes de caractéristiques extraites par le backbone

Augmentation de données À partir des images annotées dites de référence, chaque agglomérat/agrégat est extrait pour constituer une bibliothèque d'agglomérats/agrégats. La figure 2 illustre l'extraction de trois agrégats. L'augmentation de données consiste alors à simuler de nouvelles images en appliquant tout d'abord de manière aléatoire un retournement et une rotation à ces agrégats, puis en positionnant de manière aléatoire ces agrégats de particules sur différents fonds vides (avec ou sans matrice, la matrice pouvant être constituée d'autres particules telles que des particules de dioxyde de silice). Partant de ces images simulées, des transformations sont appliquées de façon aléatoire aux images pendant l'entraînement (flou gaussien, normalisation du contraste, bruit gaussien additif, multiplication de la valeur des pixels, ...).

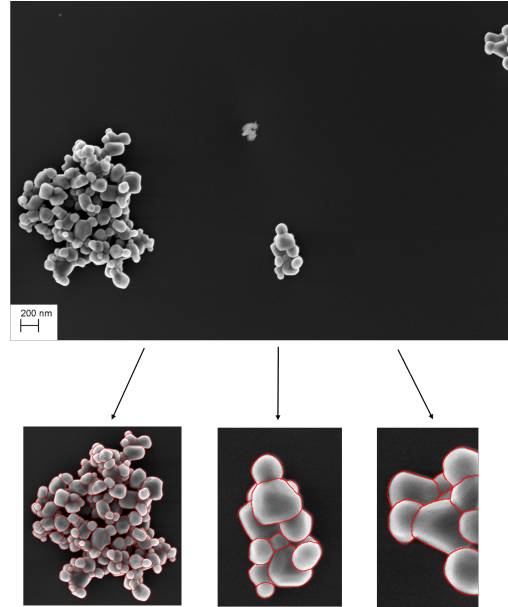


Figure 3: Image originale de particules de TiO_2 mesurée au MEB (figure du haut) et les agrégats annotés (3 figures du bas) qui seront utilisés pour simuler de nouvelles images.

Transfert d'apprentissage Le transfert d'apprentissage (Torrey et Shavlik (2010)) sert à palier au manque de données et consiste à initialiser le réseau avec les poids d'un réseau avec la même architecture mais entraîné pour une tâche différente. L'approche la plus couramment utilisée consiste à entraîner d'abord les couches finales du réseau ³, puis, après un nombre choisi d'itérations, à entraîner l'ensemble du réseau ⁴ pour la tâche spécifique, dans notre cas la segmentation des particules de TiO_2 . Les poids transférés proviennent de l'algorithme entraîné sur la base de données MS COCO proposée par Lin *et al.* (2014) qui contient 91 catégories d'objets distinctes et près de 2 500 000 instances annotées sur 328 000 images.

Ajustement des hyper-paramètres Les images MEB étudiées comportent un grand nombre de particules de TiO_2 dans chaque image, parfois plusieurs centaines, voire des milliers. Or, l'algorithme a été initialement développé pour détecter des personnes, des voitures, etc... En d'autres termes, le nombre d'instances de classes par image est drastiquement plus élevé. L'une des premières adaptations de l'algorithme a donc été réalisée pour le réseau assurant la proposition de régions (RPN). L'objectif du RPN est de parcourir toutes les cartes de caractéristiques extraites par le backbone (approche par fenêtre glissante via des éléments géométriques appelés "ancres"), d'identifier les objets et d'estimer précisément les boîtes englobantes pour chaque objet. Les modifications suivantes ont été apportées afin de prendre en compte la spécificité de la tâche de segmentation : les tailles des ancres ont été modifiées en fonction de la taille minimale et maximale des particules de TiO_2 (8, 16, 32,

³les poids transférés dans les premières couches restent alors fixes

⁴l'ensemble des poids du réseau sont mis à jour

64 et 96 pixels); le nombre d’ancres entraînées a été adapté par rapport au nombre maximal de particules dans une image (1024); l’écart entre deux ancres consécutives a été fixé à (1) en raison du phénomène d’agrégation des particules.

Entraînement du réseau de neurones Le jeu de données d’entraînement comporte 699 images, dont 77 images réelles et 622 images simulées via la stratégie d’augmentation de données présentée précédemment. La stratégie d’entraînement reste très artisanale à ce stade et a été purement guidée par les performances constatées. Les têtes du réseau ont été entraînées pendant 38 itérations⁵ avec un taux d’apprentissage de 0.001 puis 4 itérations avec un taux d’apprentissage de 0.0001. L’ensemble du réseau a ensuite été entraîné pendant 28 itérations avec un taux d’apprentissage de 0.001 et enfin 7 itérations avec un taux d’apprentissage de 0.0001. L’algorithme d’optimisation choisi est un algorithme de descente de gradient stochastique (SGD) avec un momentum (0.9) et un seuillage de la norme du gradient (5.0). Comme suggéré par Krogh et Hertz (1992), nous utilisons la régularisation L_2 avec une décroissance de poids (0.0001). Le réseau a été entraîné sur une simple carte graphique GPU NVIDIA GeForce RTX 2080 avec 8 Go de mémoire.

3 Classification de l’état d’agrégation des particules de TiO_2

Cette étape de classification fait office de sélection des données utiles pour l’estimation de la distribution en taille pour laquelle seules les particules entièrement imagées doivent être mesurées. Les particules détectées sont donc réparties au sein des 5 classes suivantes : Isolée (*la particule est complètement imagée et située en dehors d’un agrégat*), Complète (*la particule est complètement imagée et située dans un agrégat*), Complète* (*la particule est complètement imagée mais imbriquée dans une autre particule, c’est l’état intermédiaire entre “complète” et masquée*), Masquée (*la particule est en partie cachée par une autre particule*), Inutilisable (*la particule est masquée par une autre particule dont la surface visible est très réduite*). Le tableau 1 illustre les différentes classes en détaillant le nombre de particules par classe pour les jeux de données d’entraînement et de test.

La base de données La base de données a été complètement créée à partir des prédictions de l’algorithme de segmentation sur les images de référence. Des experts en nanométrie ont ensuite statué sur l’état d’agrégation de chaque particule à partir des vignettes extraites. Pour rappel, une segmentation correspond à un masque binaire. Les images d’entrée sont ici composées de deux canaux : le premier contient l’image de la particule en nuance de gris et le deuxième contient le masque de segmentation produit lors de l’étape précédente. Nous avons ajouté la classe “A jeter” pour les particules mal segmentées.

⁵le terme itération ici fait référence à un seul passage complet à travers l’ensemble des données d’entraînement lors de l’apprentissage du modèle (epoch en anglais)






Vignette	Etat d'agrégation	# (Entraînement)	# (Test)
	Isolée	567	54
	Complète	1129	98
	Complète*	2329	250
	Masquée	4611	478
	Inutilisable	1747	183
	À jeter	2789	320
	Total	13172	1383

Table 1: Les différents états d'agrégation des particules de TiO_2 et le nombre de particules présentes dans les jeux de données d'entraînement et de test pour la tâche de classification.

Le réseau de neurones sélectionné La figure 4 présente le schéma du réseau VGG16 utilisé, initialement développé par Simonyan et Zisserman (2014). La philosophie de cette architecture consiste en l'enchaînement de blocs convolutifs munis de fonction d'activation ReLU, chacun séparé par une couche de max pooling permettant de réduire la résolution spatiale. Dans une architecture VGG, lorsque la résolution spatiale est réduite par 2 (en hauteur et en largeur, soit 4 fois moins de pixels), on augmente par 2 la profondeur de l'image (2 fois plus de filtres).

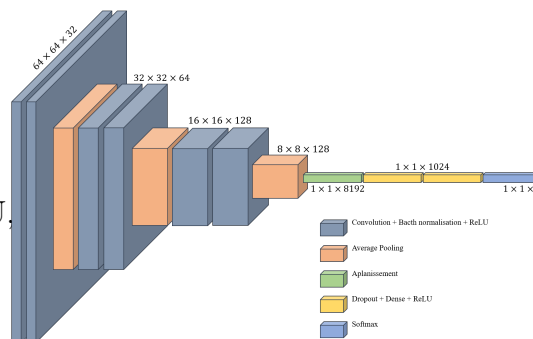


Figure 4: Architecture du réseau VGG16 utilisé pour la tâche de classification de l'état d'agrégation des particules de TiO_2 .

Les dernières couches complètement connectées, suivi d'une couche softmax (régression logistique multinomiale) servent ensuite de classifieur. L'entraînement de ce type de réseau est extrêmement classique et ne sera pas détaillé ici.

4 Résultats

Dans cette section, nous présentons les résultats obtenus respectivement par le Mask-RCNN modifié (section 2) et par le modèle VGG16 (section 3) pour les tâches de segmentation et de classification de l'état d'agrégation des particules de TiO_2 . Il est à préciser qu'il n'existe pas aujourd'hui dans la littérature d'approches concurrentes complètement automatisées auxquelles se comparer. Pour cette raison, les comparaisons numériques proposées

considèrent l’opérateur humain comme la référence.

4.1 Segmentation

Le jeu de données de test est composé de 19 images pour un nombre total de 3741 particules de TiO_2 . Ces 19 images couvrent plusieurs configurations rencontrées régulièrement en pratique : particules de TiO_2 agrégées, dispersées ou mélangées avec d’autres types de particules.

performance en détection Il s’agit simplement ici d’évaluer sur le jeu de test le nombre de particules détectées par l’algorithme de segmentation présenté en section 2 en fonction du statut d’agrégation des particules. Le tableau 2 montre que la détection des particules d’intérêt pour l’estimation de la distribution en taille est satisfaisante ($> 99\%$ pour les particules de la classe Complète et $> 96\%$ pour les particules de la classe Complète*).

	Total	Complète	Complète*	Masquée	Inutilisable
Référence	3741	341	515	1302	1583
Mesure	3135	339	495	1253	1048
% bonnes détections	83.80	99.41	96.11	96.23	66.2

Table 2: Nombre de particules détectées sur le jeu de test en fonction de l’état d’agrégation des particules.

performance en segmentation Évaluer les masques prédits pour chaque particule nécessite une mesure de similarité entre le masque de référence et le masque généré par l’algorithme. Il existe un grand nombre de mesures dans la littérature (indice de Jaccard (1901), indice de Tversky (1977), ...), nous faisons le choix du coefficient de Sørensen–Dice ou coefficient de similarité de Dice (développé indépendamment par Dice (1945) et Sorensen (1948)), il est donné par la formule suivante :

$$DSC = \frac{2 \times |A \cap B|}{|A| + |B|}$$

, où A et B désignent respectivement les masques binaires de référence et prédit et $|A|$ correspond la taille de A, autrement dit le nombre de pixels ayant pour valeur 1 dans le masque A. Une parfaite concordance entre les deux masques se traduit par un coefficient de 1, tandis qu’une valeur de 0 indique une absence de chevauchement entre les masques. Le tableau 3 présente les paramètres statistiques de la distribution du coefficient de Sørensen–Dice sur les particules détectées sur le jeu de test en fonction de l’état d’agrégation. L’analyse des résultats est similaire à celle pour la détection, c’est-à-dire que l’algorithme prédit des masques similaires aux masques de référence pour les particules d’intérêt (DSC moyen > 0.95 pour les particules des classes Complète et Complète*).

DSC

	Moyenne	Médiane	Ecart-type	Min	Max
Total	0.936	0.948	0.040	0.750	0.989
Complète	0.953	0.959	0.024	0.819	0.987
Complète*	0.953	0.958	0.022	0.858	0.986
Masquée	0.949	0.957	0.026	0.765	0.989
Inutilisable	0.903	0.913	0.046	0.756	0.975

Table 3: Statistiques du coefficient de Sørensen–Dice sur les particules détectées sur le jeu de test.

estimation du diamètre de surface équivalent

Les mesures de performance se sont intéressées jusqu’ici à la détection des particules et à la prédiction des masques de segmentation. Pour obtenir la distribution en taille d’un échantillon, l’étape finale consiste à calculer le diamètre équivalent de chaque particule. Ici nous nous intéressons au diamètre de surface équivalent. La figure 5 montre la distribution des résidus en pourcentage entre le diamètre de surface équivalent calculé à partir de la segmentation de référence et la segmentation prédite automatiquement. Les résultats révèlent que 96 % des mesures sur le jeu de test présentent moins de 5 % d’erreur sur l’estimation du diamètre et également que 51 % des mesures présentent moins de 1 % d’erreur.

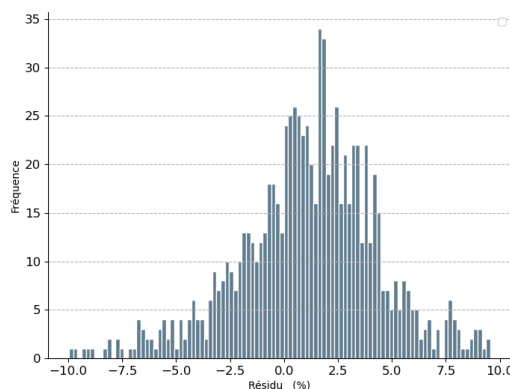


Figure 5: Histogramme des résidus (en pourcentage) pour le diamètre de surface équivalent sur le jeu de test.

4.2 Classification

Sur le jeu de test détaillé en section 3, le VGG16 utilisé atteint un score de bonne classification de 76.3% (soit 23.8% de mauvaises classifications). Bien que ce score puisse paraître assez faible, il est du même ordre de grandeur que la performance estimée sur un opérateur humain, ce qui en fait par conséquent un bon classifieur. En effet, il peut être très complexe de différencier les particules des classes masquées/inutilisables ainsi que les particules des classes complète/complète*/masquée comme indiqué par le tableau 4 présentant la matrice de confusion des prédictions.

Le tableau 5 récapitule le nombre de bonnes et mauvaises classifications en fonction du seuil choisi sur la prédiction du modèle. Bien évidemment, augmenter ce seuil conduit à réduire considérablement le nombre de particules (97.2% de bonnes détections pour un seuil de 0.95 qui conduit à conserver 394 particules pour un nombre initial de 1383, soit 28.5%

Vraie classe	Classe prédite					
	Complète	A jeter	Isolée	Masquée	Complète*	Inutilisable
Complète	77	1	0	1	19	0
A jeter	2	251	1	34	9	23
Isolée	0	2	52	0	0	0
Masquée	4	15	0	376	59	24
Complete*	30	8	0	32	177	3
Inutilisable	0	14	0	42	5	122

Table 4: Matrice de confusion de la tâche de classification par VGG16.

seulement des particules conservées). L’analyse de ces résultats conduit à sélectionner un seuil de 0.80 qui conduit à 88.7% de bonnes classifications en maintenant plus de 50 % des particules.

	Seuil						
	0.0	0.5	0.6	0.7	0.8	0.9	0.95
Complète	77 (78.6)	71 (78.0)	62 (84.9)	56 (84.5)	50 (89.3)	41 (95.3)	33 (97.1)
A jeter	251 (78.4)	242 (81.8)	234 (83.9)	221 (86.0)	203 (91.0)	181 (95.8)	160 (98.8)
Isolée	52 (96.3)	52 (96.3)	52 (96.3)	52 (96.3)	52 (96.3)	52 (96.3)	52 (96.3)
Masquée	376 (78.7)	359 (75.1)	324 (83.7)	273 (85.0)	224 (89.6)	163 (95.9)	97 (97.0)
Complete*	177 (70.8)	165 (71.7)	147 (76.2)	118 (80.3)	86 (80.4)	37 (80.4)	19 (86.4)
Inutilisable	122 (66.7)	117 (70.5)	98 (76.0)	85 (77.3)	71 (84.5)	44 (91.7)	22 (95.7)
Total	1055 (76.3)	1006 (78.3)	917 (82.3)	805 (84.6)	686 (88.7)	518 (94.4)	383 (97.2)
# particules	1383 (100.0)	1284 (92.8)	1114 (80.5)	952 (68.8)	773 (55.9)	549 (39.7)	394 (28.5)

Table 5: Nombre de bonnes classifications en fonction de la classe de la particule et du score de prédiction (les pourcentages sont en ()). # particules précise le nombre de particules restantes après seuillage.

5 Conclusion

L’automatisation de la caractérisation de la taille des particules de dioxyde de titane (TiO_2) obtenues par MEB devient souhaitable compte-tenu du temps et des ressources nécessaires à l’heure actuelle pour réaliser cette tâche, souvent réalisée manuellement, qui est rendue difficile par la non sphéricité des particules et à leur tendance à fortement s’agréger. Recourir à l’apprentissage statistique malgré une base de données métier de taille réduite est possible aujourd’hui en utilisant les techniques de transfert d’apprentissage et d’augmentation de données. Les résultats obtenus sont convaincants avec notamment 96 % des mesures sur le jeu de test présentant moins de 5 % d’erreur sur l’estimation du diamètre de surface équivalent. Ces résultats ouvrent clairement la voie à l’automatisation complète de ce type de caractérisation purement basée sur la vision par ordinateur, tâche sur laquelle ces algorithmes

ont montré leur supériorité depuis quelques années. La suite des travaux portera sur la quantification de l'incertitude associée aux prédictions des réseaux de neurones utilisés dans cette chaîne de traitements automatisés.

Bibliographie

Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology*, 26(3), pp. 297-302.

Sorensen, T. (1948). A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons. *Biologiske skrifter*, 5, pp. 1-34.

Jaccard, P. (1901). Distribution de la flore alpine dans le bassin des Dranses et dans quelques régions voisines. *Bulletin de la Société vaudoise des sciences naturelles*, 37, pp. 241-272.

Tversky, Amos (1977). Features of Similarity. *Psychological Review*. 84 (4), pp. 327–352.

Weir, A., Westerhoff, P., Fabricius, L., Hristovski, K., et Von Goetz, N. (2012). Titanium dioxide nanoparticles in food and personal care products. *Environmental science and technology*, 46(4), pp. 2242-2250.

Hwang, J. S., Yu, J., Kim, H. M., Oh, J. M., et Choi, S. J. (2019). Food additive titanium dioxide and its fate in commercial foods. *Nanomaterials*, 9(8), 1175.

Torrey, L., et Shavlik, J. (2010). Transfer learning. *In Handbook of research on Machine Learning applications and trends: algorithms, methods, and techniques*, pp. 242-264. IGI global.

Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., ... et Zitnick, C. L. (2014). Microsoft COCO: Common objects in context. *European Conference in Computer Vision–ECCV 2014*, 13, pp. 740-755.

Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., et Belongie, S. (2017). Feature pyramid networks for object detection. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2117-2125.

He, K., Gkioxari, G., Dollár, P., et Girshick, R. (2017). Mask R-CNN. *Proceedings of the IEEE international conference on computer vision*, pp. 2961-2969.

Krogh, A., et Hertz, J. (1991). A simple weight decay can improve generalization. *Advances in neural information processing systems*, 4, pp. 950-957.

Simonyan, K., et Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv:1409.1556*.